DATA-DRIVEN MODELING OF HYDROCLIMATIC TRENDS AND SOIL MOISTURE: MULTI-SCALE DATA INTEGRATION AND DECISION SUPPORT

BY

EVAN JOSEPH COOPERSMITH

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Environmental Engineering in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Barbara Minsker, Chair
Professor Murugesu Sivapalan
Professor Praveen Kumar
Assistant Professor Carl Bernacchi

ABSTRACT

The techniques and information employed for decision-making vary with the spatial and temporal scope of the assessment required. In modern agriculture, the farm owner or manager makes decisions on a day-to-day or even hour-to-hour basis for dozens of fields scattered over as much as a fifty-mile radius from some central location. Following precipitation events, land begins to dry. Land-owners and managers often trace serpentine paths of 150+ miles every morning to inspect the conditions of their various parcels. His or her objective lies in appropriate resource usage – is a given tract of land dry enough to be workable at this moment or would he or she be better served waiting patiently? Longer-term, these owners and managers decide upon which seeds will grow most effectively and which crops will make their operations profitable. At even longer temporal scales, decisions are made regarding which fields must be acquired and sold and what types of equipment will be necessary in future operations. This work develops and validates algorithms for these shorter-term decisions, along with models of national climate patterns and climate changes to enable longer-term operational planning.

A test site at the University of Illinois South Farms (Urbana, IL, USA) served as the primary location to validate machine learning algorithms, employing public sources of precipitation and potential evapotranspiration to model the wetting/drying process. In expanding such local decision support tools to locations on a national scale, one must recognize the heterogeneity of hydroclimatic and soil characteristics throughout the United States. Machine learning algorithms modeling the wetting/drying process must address this variability, and yet it is wholly impractical to construct a separate algorithm for every conceivable location. For this reason, a national hydrological classification system is presented, allowing clusters of hydroclimatic similarity to emerge naturally from annual regime curve data and facilitate the

development of cluster-specific algorithms. Given the desire to enable intelligent decision-making at any location, this classification system is developed in a manner that will allow for classification anywhere in the U.S., even in an ungauged basin. Daily time series data from 428 catchments in the MOPEX database are analyzed to produce an empirical classification tree, partitioning the United States into regions of hydroclimatic similarity. In constructing a classification tree based upon 55 years of data, it is important to recognize the non-stationary nature of climate data. The shifts in climatic regimes will cause certain locations to shift their ultimate position within the classification tree, requiring decision-makers to alter land usage, farming practices, and equipment needs, and algorithms to adjust accordingly. This work adapts the classification model to address the issue of regime shifts over larger temporal scales and suggests how land-usage and farming protocol may vary from hydroclimatic shifts in decades to come.

Finally, the generalizability of the hydroclimatic classification system is tested with a physically-based soil moisture model calibrated at several locations throughout the continental United States. The soil moisture model is calibrated at a given site and then applied with the same parameters at other sites within and outside the same hydroclimatic class. The model's performance deteriorates minimally if the calibration and validation location are within the same hydroclimatic class, but deteriorates significantly if the calibration and validates sites are located in different hydroclimatic classes. These soil moisture estimates at the field scale are then further refined by the introduction of LiDAR elevation data, distinguishing faster-drying peaks and ridges from slower-drying valleys. The inclusion of LiDAR enabled multiple locations within the same field to be predicted accurately despite non-identical topography. This cross-

application of parametric calibrations and LiDAR-driven disaggregation facilitates decision-support at locations without proximally-located soil moisture sensors.

To Mom, Dad, and Rebecca

ACKNOWLEDGEMENTS

The production of a dissertation is a life-consuming enterprise that threatens one's sanity and repurposes one's young adulthood. That I have escaped with some modicum of sanity and joie de vivre, is, in no small part due to the love and support from family, from friends, and of course, from my wife.

To Barbara Minsker, my adviser, I thank you for recognizing my capacity as a scholar, my love of teaching, and my independent streak as a human being…and nurturing them all.

To Murugesu Sivapalan, who taught a course in hydroclimatolgy whose final project ultimately became the framework for multiple publications and this doctoral work. I was fortunate to find myself in the right place at the right time.

To Andrew Collier, whose friendship during a psychologically-grueling first year facilitated my finding myself romantically and academically, I owe you debt of gratitude I can never repay.

To the folks at John Deere's Technological Innovation Center (JDTIC), who not only provided the funding for the majority of the work herein, but also offered their research park offices as a scholarly oasis, I thank you for offering me your friendship and acceptance, but also for the corporate perspective I had lacked in academia.

To Mikko Tuomela, at JDTIC, who helped me acquire and process massive quantities of LiDAR data – I do hope our collaborations do not end here.

To team Disc in a Box, who reminded me that graduate school is more fun when muddy.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND BACKGROUND LITERATURE

This chapter will introduce the three objectives of this research, offering an overview of the expectations and approaches to each. Following the brief explanation of each substantive question to be answered, this chapter will present the majority of the literature that forms the basis of this research. The chapter begins with the wetting/drying modeling, continues into the process and literature of hydroclimatic classification, and then finally, presents an approach to and some of the literature of hydroclimatic change. Further discussion then addresses some of the issues of integrating data at a variety of spatial and temporal scales. This section will introduce each objective followed by the relevant research, beginning with wetting/drying, proceeding to climate classification, continuing with regional analysis of hydroclimatic change, and concluding with discussion of cross-application of models developed at multiple locations.

1.1 Decision Support for Wetting/Drying

The hydrologic processes of wetting and drying play a crucial role in construction, agriculture, and back-country studies involving heavy equipment on unpaved terrain. When soil conditions moisten, equipment can become mired, causing expensive delays. While experienced users may assess soil conditions before entering off-road areas, novice users or those who must remotely assess sites before traveling may have difficulty assessing conditions reliably. For example, farmers can trace serpentine paths of over one hundred miles *daily* to assess the status of land spanning fifty miles in each direction. While the definition of "too wet" varies with usage, one means of assessing dryness is remotely-monitored *in situ* sensors. Unfortunately,

land owners hesitate to place sensors due to monetary costs, complexity, and sometimes infeasibility of physical visits to remote locations. Chapter II addresses these limitations by modeling the wetting/drying process through machine learning algorithms fed by hydrologic data - remotely assessing soil conditions using only publicly-accessible information. In conjunction with enhanced national climate classification, such models will be applicable at any location nationally.

Previous work has made forays into soil drying assessments over a diverse set of geographic locations, climate conditions, and functional objectives. The primary dynamic process affecting soil drying is precipitation (Entekhabi and Rodriguez-Iturbe, 1994). For this reason, models of wetting and drying have often focused upon an "antecedent precipitation index" (API), using a pre-set window of previous rainfall to estimate current levels of soil moisture (Saxton and Lenz, 1967). This particular concept of calculating an API has been applied in a variety of contexts: in conjunction with microwave sensing for soil moisture estimation (Blanchard et al, 1981), soil water recession modeling for agriculture (Choudhury and Blanchard, 1983), and for weather prediction (Wetzel and Chang, 1988). Another approach is the development of a stochastic model to estimate soil moisture distributions using daily rainfall and an initialization of soil moisture values (Farago, 1985). However, both the API and stochastic approach require an initial condition for soil moisture at the location where estimates are desired. This hampers applicability at many locations that do not have soil moisture sensors.

Other models have taken a hydrologic approach, employing precipitation and surface radiation to estimate soil moisture (Capehart and Carlson, 1994), but these models require boundary conditions, initial conditions, and parameters of a thermal and/or hydraulic nature that can be difficult to obtain broadly. Pan et al. (2003, 2012) addressed this concern by deriving a

"diagnostic soil moisture equation" from a stochastic, linear partial differential equation. Soil moisture then becomes a function of a temporally-decaying sum of previous rainfall. Their approach no longer requires an initial condition, nor recalibration, but does require a soil moisture sensor at the location in question to calibrate the equation initially. Measuring soil moisture directly is plausible, but soil heterogeneity may necessitate numerous sensors to address spatial variation of soil moisture adequately (Pan and Peters-Lidard, 2008). The alternative approach of a soil water balance can be applied, but must be recalibrated frequently, since errors are cumulative (Jones 2004).

In the agricultural arena, Gamache et al (2009) developed a soil drying model, but its predictions require data from cone penetrometer and soil moisture sensors, two data sources that are not currently available at most remote sites. Another inquiry along similar lines uses knowledge of soil types, which is theoretically public, but then continues to require soil moisture levels from proximal sensors (Chico-Santamaria et al, 2009).

Other approaches eschew the strategic placement of soil sensors in favor of modeling tire slip as a function of tractor properties (Sahu and Raheman, 2008) or other details such as vehicle type, speed, load distribution, number of passes, etc (Pytka, 2009, and Lamande and Schjonning. 2008). These vehicle-specific properties are often unavailable outside of research studies. Another early work attempts to assess the suitability of site conditions, but uses very specific information that is not likely to be available to most applications, such as "stress-strain rate relations" or the results of triaxial tests (Sharifat and Kushwaha, 2000).

Lee and Wang (2009) focus solely on radar and other remote sensing data, but consider only the properties of snow coverage along with the hardness and density of mixtures of snow and ice. These traits are not appropriate for warmer weather conditions of interest to agriculture.

Silva et al. (2009) modeled drying properties of sugarcane fields in Brazil, but their model requires that the prediction occur in a well-specified location within a pre-determined time frame.

Alternatively, considerable prior research seeks techniques for improving agricultural conditions (Tullberg et al, 2007; Shoop et al, 2002; and Lebert et al, 2006) or minimizing the effects of traffic (Raper, 2005) rather than delivering a dryness assessment. For instance, one paper recommends a protocol for improving drying via the application of manure (Mosaddeghi et al, 2000).

1.2 National Scale Hydroclimatic Classification

This section is aimed at developing a catchment classification system that will help organize a large and diverse population of catchments within the continental United States into homogeneous groups on the basis of climate seasonality and runoff regime. The work is part of a broader study aimed at better understanding of the physical controls of the Flow Duration Curve (FDC). It has been motivated by the observation that a catchment's regime curve (ensemble mean of the within-year variation of runoff) has a major impact on the shape of the FDC (Yokoo and Sivapalan, 2011), thus serving as the connective tissue between high and low flows that appear at the extreme ends of the FDC. This connection is formalized by developing a climatic classification system, based upon regime curves, that incorporates hydrologic information.

Through numerical simulations with a physically-based rainfall-runoff model applied to hypothetical catchments, Yokoo and Sivapalan (2011) showed that the FDC of total runoff can be partitioned into two components, i.e., the FDC of the surface (or fast) flow and that of subsurface (or slow) flow. This result has been further confirmed by the comprehensive analysis

of the FDCs of some 200 catchments located within the continental United States by Cheng et al. (2012). Yokoo and Sivapalan (2011) further argued that while both the fast and slow flow components are driven by different climate and landscape properties, the FDC of the slow, subsurface flow component closely resembles and could be more easily reproduced from the catchment's regime curve. If this is true, then spatial variations of the regime curve, and associated climatic and landscape controls that result from their interactions, could help explain the regional patterns of the FDCs within the continental United States. So while understanding of the process controls of the regime behavior is important in its own right, it is also valuable for understanding the controls of the FDC.

Catchments everywhere are highly variable, displaying enormous complexity, with a large number of degrees of freedom, which makes it very difficult to make general statements about their responses. Yet, despite substantial heterogeneity and the complexity of their responses exhibited by observations, experience with modeling studies and predictions indicates that at the catchment scale simple models with a small number of parameters can describe the majority of catchment responses (Sivapalan, 2003). This has encouraged hydrologists to organize and classify catchments into homogeneous or similar groups on the basis of a small number of explanatory variables, as a vehicle towards generating improved understanding and predictions (Dooge, 1986; Blöschl and Sivapalan, 1995; McDonnell and Woods, 2003, Olden et al, 2011). Due to the self-organization of climatic and landscape features arising from their co-evolution, and their impact on multi-scale process interactions and feedbacks, any catchment classification system must be necessarily holistic.

One of the pivotal differences between this work and its predecessors is the scope of the classification attempts. For instance, a finely-detailed study by Mosley (1981) classified

hydrologic responses in 175 small catchments in New Zealand, resulting in narrowly-defined characteristics and finely split classes. Ogunkoya (1988) classified 15 catchments in Nigeria, but considered lithographic details and other features that may be less appropriate if a classification scheme is to be broadly applied and minimalist in its information requirements, as is the objective of this analysis. Burn (1997) applied seasonality metrics to help understand flood frequencies in 59 prairie catchments in central/western Canada chosen specifically because they experience comparable climates, and thus all present hydrologic regimes driven by flood events from spring snowmelt. Their results, while useful, do not address the tremendous climatic diversity that can occur at the continental scale. Recognizing this, Burn and Goel (2000) chose a more diverse assortment of catchments in India, using a k-means technique to effectively extract groups of similar catchments. While these catchments exhibited more geographic complexity than the previously discussed studies, this location is still somewhat hydrologically limited. In addition, clustering algorithms of this kind present groups that are similar without specifying the physical drivers that contribute to such similarity - an imperative for deeper understanding of process controls.

Rather than simply examine how quantitative characteristics of catchments in various regions are optimally organized, this analysis also focuses upon why these catchments present the observed climatic and hydrologic characteristics that they do. As mentioned earlier, to understand the physical controls on the FDC, one can classify runoff regimes using empirical runoff regime data, as seen in Haines et al (1988), where clusters of catchments with similar flow regimes were obtained by minimizing within-group variance of clusters of monthly streamflows. While this procedure does yield qualitative explanations, they were generated after the fact, rather than as part of the analysis itself. Qualitative insights are strongest as the result of

objective, rather than reflective analysis. One way of gaining qualitative insights from an objective process is the use of hydrologic signatures. Wagener et al. (2007) proposed a classification system that is based on similarity of key signatures of catchment runoff response, including, with decreasing timescale, inter-annual variability, regime curve (i.e., mean within-year variability of runoff), and the flow duration curve (FDC). Taking this idea further, Sawicz et al (2011) classified catchments located in the eastern half of the United States, using several catchment-based signatures including the runoff ratio, the slope of a flow duration curve, and other streamflow properties. This was followed by a comparative study of several catchments based on detailed physically-based modeling that can account for differences in topography, soil types, geomorphology, and vegetation (Carillo et al., 2011). These studies began investigating the physical underpinnings of the groups that emerge from classification – we intend to continue in a similar vein, using simple regime-curve-based features.

With respect to hydrologic, signature-based classification, there has been considerable success in developing similarity measures and catchment classification on the basis of mean annual runoff, expressed in terms of the Budyko curve and the aridity index (Budyko, 1974; Zhang et al., 2001). The focus on the regime curve in this chapter is a natural extension to establish the basis for similarity of catchment responses. Whereas the competition between water available and energy available governs similarity at the annual timescale, the shape of the regime curve is governed additionally by the relative timing of precipitation and potential evaporation, and the ability of the landscape to store and release water.

Frameworks for climate classification were first applied broadly via the Koppen-Geiger system - identifying similar climates using basic information on the variability of precipitation and temperature (Koppen and Geiger, 1936), and later updated by Peel et al. (2007). The

classification of regime behavior presented in this chapter can be seen as a precursor to a possible hydrological extension of the Koppen-Geiger system towards classification of catchment responses. The Koppen-Geiger system is based on the number of months in which average precipitation or average temperature exceeds a given threshold. However, by excluding hydrology from the system, it fails to distinguish certain catchments that display different filtering behavior. Consider that Koppen-Geiger classifies the entire south-eastern United States identically. Understanding the distinctions in rainfall/runoff timing allows for more nuanced analysis of the FDC – this was the hypothesis raised by Yokoo and Sivapalan (2011) upon which this chapter builds.

Regime curves, in addition to their hydrological importance discussed above, present an image that is meaningful to decision-makers. The growing season is visible in terms of the months in which precipitation and runoff work in concert in a manner conducive to agriculture. This, in conjunction with the limited data demands, makes the regime curve an ideal basis for classification. However, the construction of these regime curves assumes stationary time series for precipitation and runoff, which, under climate change, are no longer valid. Stated succinctly, "stationarity is dead." (Milly et al, 2008) For this reason, the next section will address the need to account for shifts in hydroclimatic regimes in classifying watersheds.

1.3 Shifts in Hydroclimatic Classification

The hydrologic classification work described in Chapter III implies that the time series data that forms the basis of those classes arrives from stationary distributions. As the literature presented in this section suggests otherwise, it is important to understand how the national hydroclimate has varied by region in the past several decades, in the hopes of better

understanding how algorithms will need to be altered to address shifts in hydroclimatic features and land use.

Due to the diversity of catchments throughout the United States, the effects of hydroclimatic shifts are not uniform. Understanding the diversity of these impacts begins with an understanding of the diversity of catchments within the United States. Within the continental U.S., catchments present considerable variability and numerous degrees of freedom in terms of their response to hydroclimatic inputs. This heterogeneity of catchments and complexity of responses notwithstanding, existing research demonstrates that, at the catchment scale, many catchment responses are well-described by comparatively simple models with manageable numbers of parameters. With this notion in mind, hydrologists have attempted to classify catchments into similar clusters based on a few key features to enhance understanding of hydrologic processes and ultimately lead to more accurate predictions (Dooge, 1986; Blöschl and Sivapalan, 1995; McDonnell and Woods, 2003; Wagener et al, 2007; Coopersmith et al, 2012).

In terms of global climate impacts during the same period, Munson et al (2012) describe climatic shifts in the southwestern regions of the United States. For land-use decision-making, this hydroclimatic variability may manifest as a change in the selection of crops (as mentioned previously), variations in yields of the same crops over a series of growing seasons (Urban et al, 2012; Lobell et al, 2013), or a shift in the timing of spring's onset (Cayan et al, 2001). In cooler regions, like New England and the Pacific Northwest, regional hydroclimatic shifts differ from the temperate and arid climates of the southern states. Snow water volume has decreased atop mountain peaks in western states (Nolin, 2012), displaying climatic effects as a function of elevation (Weare and Blossier, 2012) with a variety of possible impacts (Beniston, 2003). In New England, warming climates are thought to have caused northern hardwood forests to shift

upslope between 1964 and 2004 (Beckage et al, 2008); future climatic shifts will likely yield a continuation of such changes, altering the composition and distribution of these northeastern forests (Evans and Perschel, 2009). In some of these cases, the distribution of vegetation has not reached equilibrium with changing climate conditions (Woodward and Beerling, 1997). It is thought to be climate impacts that have driven the uphill movement of the deciduous forests to replace evergreens (Tang et al, 2012).

While fully-detailed projections of climates in future years and fully-specified depictions of previous changes are highly complex and best left to climate models, many of which offer widely varying regional predictions (Cai et al., 2009), the classification system by Coopersmith et al. (2012) described in chapter III can offer simpler qualitative and quantitative insights to inform modeling and decision-making. For example, in certain regions, the growing season that shapes land-use decisions may lengthen or shorten, altering the availability and productivity of arable land (Cai et al., 2011; Zhang and Cai, 2011). Subsequently, crop yields and timing are subject to change, or in more extreme cases, choosing a different crop altogether may be worth consideration.

Having developed a framework for classifying locations of similar hydroclimatic properties in chapter III, then analyzed how these regions might have varied over the past several decades in chapter IV, the question can then be asked – if a model is calibrated at one location, is it applicable elsewhere within that same class?

CHAPTER 2

DECISION-SUPPORT FOR WETNESS/DRYNESS MODELING

2.1 Problem Definition and Overview

For the purposes of this hydrological analysis, precipitation and potential evaporation data serve as inputs into machine learning algorithms that output an estimate of soil dryness conditions This notion of dryness represents a user-defined assessment with qualitatively consistent designations for a particular application. For example, agricultural applications may possess very different notions of acceptable soil conditions, but provided the algorithm is given training data consistent to one particular context, it will adapt appropriately. This current analysis focuses upon a test case of agricultural soil drying, where "dry" implies that a given tract of farmland is viable for a particular type of work (e.g., planting, crop treatment, or harvesting) on a given day.

The objective of this work is to present a proof of concept that machine learning tools, even with limited, exclusively public data, can generate sufficiently accurate "nowcasts" of wetting/drying, using a test site located at the University of Illinois' South Farms site in Urbana, IL. These nowcasts will be compared against qualitative assessments gathered over the growing season of 2010.

This section begins in with a presentation of the case study, a brief overview of the geography and relevant features of the South Farms test site that is the focus of the data analysis presented in this work. The next segment describes the methodology used in remotely estimating dryness from public data sources. Continuing the results of the case study application are presented and the relative performance of the algorithms are compared. Finally an

assessment of which machine learning techniques have shown themselves to perform most successfully is made, followed by a brief discussion of potential future enhancements and other applications of this research.

2.2 Case Study – South Farms, Urbana, IL

The methods developed in this study are tested at the University of Illinois South Farms located in Urbana IL, which is classified as a continental or microthermal climate, Dfa by the Koppen-Geiger classification system (Koppen, 1936; updated by Peel et al, 2007). The specific climate zone is characterized by a warm, humid summer and colder, drier winters. Annual rainfall levels, gathered from 1990-2011 at the Illinois Climate Network (ICN) sensor located near the test site (figure 2.1), average approximately 1,013 mm per year. The potential evapotranspiration estimate over the same time period is 1,046 mm per year. The warmest month is July and the coolest is January, with average daily temperatures of 75.1 and 26.9 degrees Fahrenheit (24.0 and -2.8 Celsius) respectively. As precipitation levels and potential evapotranspiration are both highest during the summer (the middle of growing season), the flat landscape yields a test site that will be characterized by multiple periods of wetting and drying during any growing season. Fields in this region are often tile-drained, which results in a shorter soil drying system memory than similar locations without the tile drains.

Figure 2.1 illustrates the location of other elements within and near the test site, which is at the Energy and Biosciences Institute (EBI) energy farm. The ICN sensors provide readings of potential evaporation (which incorporates solar radiation, humidity, wind, temperature, etc) and precipitation. To the southeast are the largest plots maintained by EBI, upon which soil condition assessments were gathered.

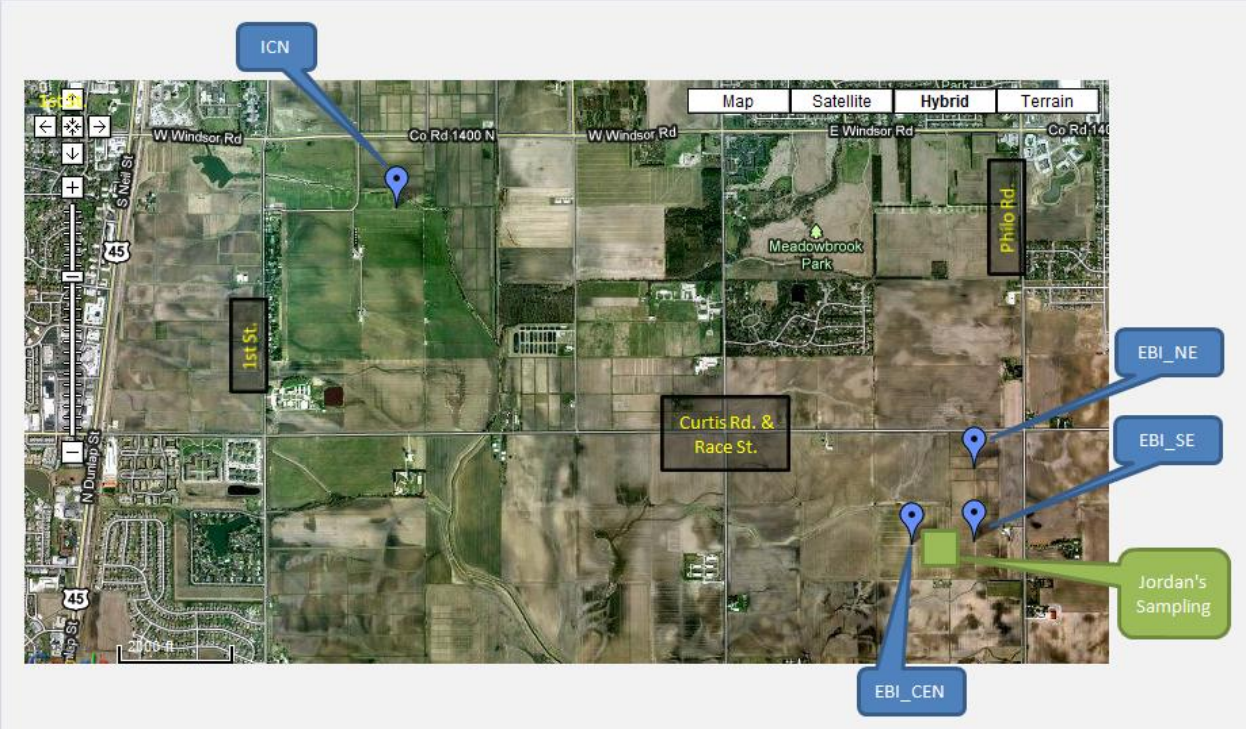


Figure 2.1 – The South Farms, Urbana-Champaign, IL

Figure 2.2 – Soil Sensor Locations

Figure 2.2, an enlarged section of the southeastern region of Figure 2.1, illustrates the locations of sixteen soil moisture sensors located on the smaller, rectangular EBI plots along with the four sensors located on the larger plots. A John Deere intern, Jordan Pitcher, provided assessments of soil conditions throughout the growing season within the green square. Mr. Pitcher has extensive agricultural experience and his assessments served as the soil dryness training and validation data for the machine learning algorithms. The sensors labeled "EBI" provide precipitation information. Data from the soil moisture sensors, though obviously inconsistent with the objective of forecasts using only public data, are included so that algorithms can be constructed with and without these data to verify that nearly equivalent accuracy can be obtained in their absence.

2.3 Methodology

This section describes the framework developed for assessing dryness based on soil drying. An overview of the approach is first provided, followed by a discussion of the various input data sources, a description of the algorithms used to assess soil drying and their outputs, and concluding with the computational tools and requirements for implementation.

This work addresses the research gaps discussed in the literature review by developing and testing a machine learning model of soil drying that requires only precipitation and potential evaporatranspiration estimates. Precipitation is widely available at high temporal resolution on a 1km by 1km grid from NEXRAD throughout the continental United States. Potential evapotranspiration is available publically in Illinois from the Illinois State Water Survey, and can be estimated in other locations using three approaches (Jensen et al 1990). The first method requires only air temperature and day length (Thornthwaite, 1948; Harmon, 1963). The second

method requires air temperature and net radiation (Priestley and Taylor, 1972). The third, and most detailed approach, requires the information from the second as well as wind speed and relative humidity (Monteith, 1965). The Illinois Climate Network data used in this analysis employs the third approach, but one of these three approaches should be applicable anywhere throughout the United States.

Adding to the challenge of decision-support modeling, the drivers of the wetting/drying process vary spatially throughout the United States. For instance, in the (largely flat) tile-drained farmlands of Illinois, drainage is extremely efficient. Thus, system memory with respect to the arrival of precipitation is generally less than 48 hours. As a result, a machine learning model accounting for the previous two days' rainfall and the available energy with which to evaporate moisture is sufficient for modeling the wetting/drying process with a high rate of accuracy. However, these independent variables may prove inadequate elsewhere. To enhance the applicability of these decision-support models on a national scale, it is necessary to develop a framework in which to define two locations as "similar" – meaning that the same machine learning model is likely to apply in both locations. In this vein, an economy of models can be constructed and calibrated, then applied where their insights are valid. This is achieved by a national hydroclimatic classification system, discussed in the subsequent section.

The first approach, the k-nearest-neighbor (KNN) algorithm, which was introduced by Fix & Hodges (1951) and deployed in many water resources and hydroinformatics applications (e.g., Kumar et al 2006, Meliker et al, 2008, McRoberts et al, 2007, Nemes et al, 2008, and Coopersmith et al, 2011), is an intuitively satisfying approach for classification, analysis, and forecasting. The algorithm simply uses current conditions to locate the most similar examples

from historical data (whose soil conditions are known) and, in turn, leverages those similar examples to estimate the current conditions.

The second algorithm, regression trees (also referred to as classification or decision trees), are non-parametric classification tools that divide data by maximum information gain (Brieman et al, 1984) and are available in most statistical programming packages (Brieman et al, 1993). This technique has been deployed in a variety of environmental contexts, such as sustainable forest resource management (Aertsen et al, 2011), crop identification for soil management (Pena-Barragan et al, 2011), and image classification for mapping the vegetation across arid rangelands (Lailiberte et al, 2007) – an enhancement of earlier decision tree work on land cover (Brodley & Freidl, 1997). Tree-based models have even been constructed in the field of finance to predict the failures of business ventures (Li et al, 2010).

The final algorithm, boosted perceptron, builds from the perceptron – the simplest, single-layer, feed-forward form of an artificial neural network (Russell & Norvig, 2010, p.729). The boosting process, using the adaboost algorithm (Freund & Schapire, 1997, described by Russel & Norvig, 2010, p. 751), has been utilized to solve a variety of environmental problems, from predicting fishery catches (Li et al, 2011), to aiding forest managers map those locations at greatest risk (Haywood & Stone, 2011) and, like classification trees, for classification of land cover from remote imagery (Stavrakoudis et al, 2011).

The soil drying assessment methodology is summarized in Figure 2.3. On the top row, four data sources are presented, including training and validation data from volunteers and the precipitation and potential evaporation inputs which define climatic conditions. Next, these data are fetched, stored, and ultimately assimilated and formatted as input streams for the three machine learning algorithms shown in the rounded rectangles at the bottom of figure 2.3. All

three algorithms generate outputs which estimate dryness either as a binary classification ("dry" or "wet") or as a probability of a given classification.

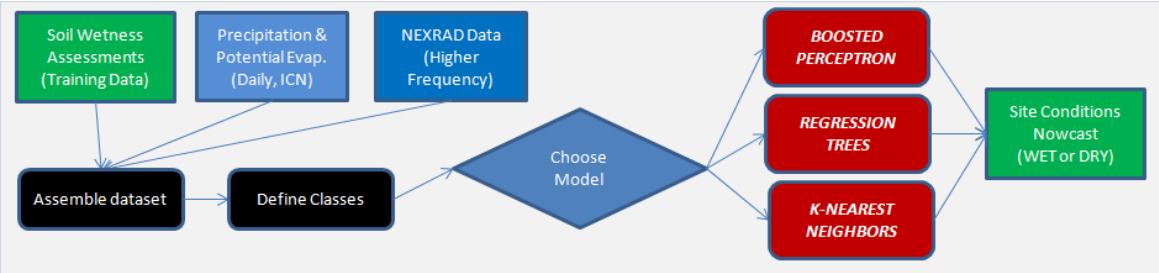


Figure 2.3 – Flow Chart of the Approach

2.3.1 Data Sources

The first source of input data is Nexrad radar data (the third box from the left, top row, figure 2.3). This high-frequency radar allows for precipitation readings hourly with nearly complete national coverage at approximately 1km x 1km granularity. The second data source, from the Illinois Climate Network, provides potential evaporation at each of their 20 sensors throughout Illinois.

The Illinois Climate Network (ICN) sensor provides hourly readings of potential evapotranspiration rates using the modified Penman-Monteith equation (Monteith, 1965). This calculation combines latent heat of evaporation, net irradiance, and constant values for specific heat of air, specific humidity, dry air density, conductivity of air, conductivity of stoma, and a psychrometric constant. The ICN sensor (see figure 2.1) currently represents the sole means of assessing the drying process using only public data. Finally, the values of soil dryness assessments, illustrated by the upper-left box of figure 2.3, are the soil condition data for model training (fitting the model parameters) and testing (validation on data that were not used for

training) given the current meteorological conditions and those from the most recent days. This qualitative metric is defined on an integral scale from 1 to 5, with the following description:

1 – The site is impassable; the equipment used might literally become stuck in the mud.

3 – The site is usable, but the equipment would leave deep ruts.

5 – The site is dry, the ground is hard, the equipment will leave only shallow track.

The values of 2 and 4 allow for descriptions which fill the gray areas between 1 and 3 or 3 and 5 respectively. As a means of developing a binary classification, a rating of 1 or 2 will be considered "wet," while a rating of 3, 4, or 5 will be considered "dry." Before settling upon this human-based, and therefore inherently subjective measurement, an alternatives was evaluated. A cone penetrometer, a tool which measures the quantity of pressure in pounds per square inch needed to compress the soil a given distance, was used in conjunction with the qualitative assessments. While its readings do bear some non-trivial correlation to the qualitative dryness metric, it is not reliably consistent with the expert assessment of whether or not the site conditions were appropriate for use at any given time.

Though naturally, a site with zero moisture that is nearly incompressible is "dry" and an extremely soft, wet soil is "wet," this distinction was found to be insufficient for assessing drying. When objective cutoffs for penetrometer readings were determined such that the proportion of "dry" days as chosen by the expert was equal to the proportion chosen by the penetrometer, disagreement occurred on over 30% of the days measured. Simply put, soil drying is a phenomenon that is not readily assessed by a single physical measurement.

2.3.2 Machine Learning Algorithms

This section will introduce the three machine learning algorithms deployed for assessing soil dryness conditions. The first subsection will address classification trees, the second will discuss k-nearest-neighbor algorithms, the third will illustrate boosted perceptrons, and the final subsection will present the techniques used to apply these algorithms to soil drying assessments.

2.3.2.1 Classification Trees

Classification trees, specifically the implementation most commonly available (Breiman et al, 1993), function via the iterative dichotimiser algorithm, ID3 (Quinlan, 1986). Predicated upon Occam's razor, which stipules that a simple theory is generally preferable to a more complex one, a classification tree aims to minimize the error associated with a prediction by splitting at each node based upon maximum information gain. A simplified example is presented in figure 2.4, where eight hypothetical drying scenarios are classified either as dry or not (labeled "YES" or "NO"). The values of each independent variable are presented alongside the dryness classifications in the boxes. "Precip1" and "Precip2" represent rainfall over the most recent 24 hours and from 24-48 hours, respectively.

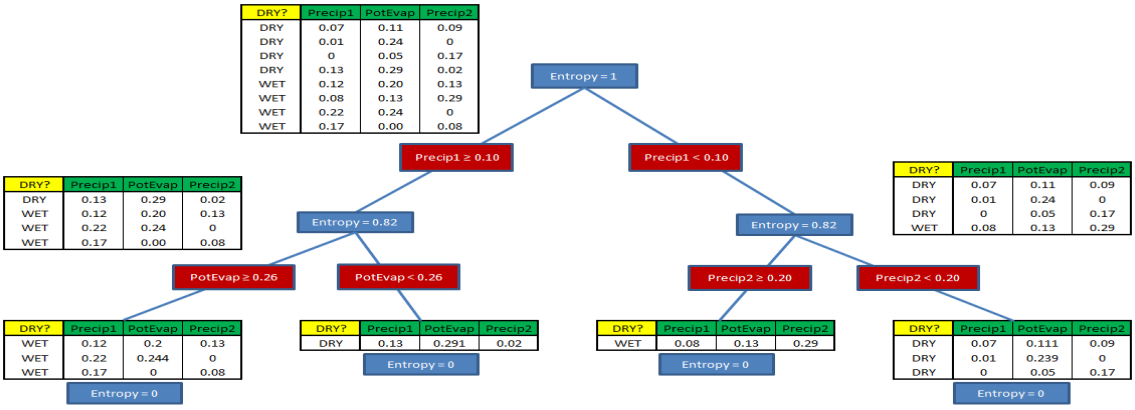


Figure 2.4 – Sample Classification Tree

Figure 2.4 illustrates that initially, there are four days classified as dry and four that are not (thus a blind guess at any given day's classification would have a 50% chance of being correct). By dividing the dataset into days with $Precip1 \geq 0.10$ and days with $Precip1 < 0.10$, two groups of four emerge, each of which have three days of one classification and a fourth of the other. This is "less disordered." A prediction based only on this single statistic, precipitation from the last 24 hours, would classify six of the eight days (75%) correctly. Finally, the two groups of four are split on potential evapotranspiration and Precip2 to yield four fully-organized clusters.

In the specific case of soil drying, consider a dataset of size *n*, in which *x* examples are "dry" and *n* – *x* examples, therefore, are "wet." Let us define information gain in terms of decreased entropy, where entropy is defined as follows:

$$E(S) = -\sum_{i=1}^{n} p_i * log_2(p_i) \qquad \text{(Equation 2.1)}$$

where $E(S)$ represents the information entropy of set *S*, n indicates the total number of examples in *S*, and p_i signifies the proportion of examples which meet criteria *i*. Inspection reveals that for perfect classification, in which all elements in *S* are of a single class, entropy is equal to zero. Conversely, for maximum entropy, in which each class *i* is equally represented within *S*, entropy is equal to unity. Thus, for the specific case of soil drying:

$$E(S) = -\left[\frac{x}{n} log_2\left(\frac{x}{n}\right) + \frac{n-x}{n} log_2\left(\frac{n-x}{n}\right)\right] \qquad \text{(Equation 2.2)}$$

Once entropy is computed for the entire dataset, an attribute *A* is chosen, such that information gain in terms of entropy is maximized:

$$G(S, A) = E(S) - \sum_{j=1}^{m} p_j(A_j) E(S_{A_j}) \quad \text{(Equation 2.3)}$$

where $G(S, A)$ is the information gain in set *S* (initially the entire data set) after splitting the dataset using attribute *A*, *m* represents the number of different possible segments of the continuous values of attribute *A* present in set *S* (for all trees discussed in this analysis, $m = 2$, as each tree is split into exactly two branches at each node), A_j refers to all values of attribute *A* that fall within segment *j*, $p_j(A_j)$ indicates the proportion of examples with attribute A_j in *S*, and S_{A_j} is simply the subset of *S* containing only examples characterized by attribute A_j.

After the first split of the dataset, further splits (nodes) are added iteratively, maximizing information gain at each step (greedy) and stopping when the number of constituent members of each leaf falls below a user-defined value deemed too low for further splitting. Alternatively, splitting can also terminate if no attribute exists such that a positive information gain can be achieved.

The advantage of classification trees, especially in circumstances where the number of examples within the training set is limited, is their ability to ignore certain attributes in situations where their information is irrelevant. Where many machine learning algorithms must utilize every feature at their disposal for each example presented, a classification tree can determine that, given one feature, the others need not be considered. Consider the case in which an intense rain event has occurred over the last six hours. Clearly, the field will be soaking wet, and unfit for use. The precipitation data from two days previously and any notion of yesterday's potential

evapotranspiration rate becomes extraneous. Thus, one branch of the classification tree may require only one split from the root node, yielding a classification instantly under certain conditions, as shown in Figure 2.5.

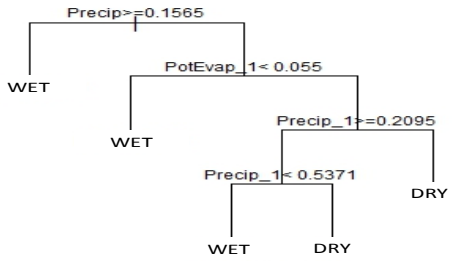


Figure 2.5 – Example of an efficient classification tree, considering only relevant features. Note that if Precipitation exceeds 0.1565, it is the only variable considered.

In the tree above, if recent precipitation exceeds a certain value, the field is classified as wet without even considering potential evapotranspiration or earlier days' rainfall. This allows the power of the algorithm and the information from the available training data to be focused upon the conditions in which soil conditions are most uncertain. With limited data, adding unnecessary features into the decision process may introduce noise without helping discern true signals. Classification trees avoid this concern, allowing for a larger number of decision variables to enter the model, while only those that conditionally prove most influential are employed for prediction.

2.3.2.2 K-Nearest-Neighbors (KNN)

The k-nearest-neighbor algorithm (KNN) searches a database for historical examples that are most similar to current conditions and then determines the proportion of those historical examples that possess the property under examination for prediction (in this case, a particular

dryness rating). Generally, similarity is determined using a simple Euclidian distance function in attribute space. Consider an input vector as follows:

$$X = (x_1, x_2, x_3, \dots, x_n) \quad \text{(Equation 2.4)}$$

Aligning the scales of each independent variable's distribution can be achieved by simply applying the following transformation of each non-normalized variable, z_i:

$$x_i = \frac{z_i - \mu_i}{\sigma_i} \quad \text{(Equation 2.5)}$$

Where μ_i represents the mean of the distribution of variable i, and σ_i denotes the standard deviation with respect to variable i. Next, let us consider a training example Y, replete with the same features as X, also normalized by the same distributions for all variables i:

$$Y = (y_1, y_2, y_3, \dots, y_n),\ y_i \sim N(0,1),\ \forall_i. \quad \text{(Equation 2.6)}$$

The n-dimensional distance function (Kumar et al, 2006, p.394) can be applied between vectors X and Y:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2},\ \forall_i \quad \text{(Equation 2.7)}$$

Thus, for a given set of current conditions X, a distance value can be determined between X and every element of the historical database. Next, a simple sort algorithm is performed, and the k

historical examples with the lowest distance values *d* are selected as a 'similar set.' At this point, in the case of soil conditions, the proportion of examples contained in the similar set which were classified as "dry" becomes the best estimate of the probability of that classification for the current situation. Essentially, KNN asks the question "what happened historically when conditions looked similar to the way they do today?"

While elegant and satisfying, the curse of dimensionality becomes unavoidable with an algorithm of this nature as more features are added, especially when the number of historical training and testing examples is low (on the order of 10^2). With each additional feature of comparison, another independent variable dimension appears, and it becomes exponentially more difficult to locate similar matches. This requires that the value chosen for k (user-selected) must decrease exponentially with each added dimension. That is, with each added dimension, either many fewer matches are considered "similar" and/or the standard of similarity becomes decidedly more lax.

To illustrate this point, consider field conditions in which no rain has fallen in the past 48 hours. If the algorithm searches for, say, ten similar days historically, it is likely that there will be ten days for which little or no rain has fallen for 48 hours. However, when we add another dimension, such as a high level of potential evaporation, finding "similar" examples becomes more challenging. Some of the previously located ten similar examples will have higher rates of potential evaporation while others will be lower. Adding the dimension of potential evaporation results in fewer similar matches (choose from the previous ten similar matches the examples with high rates of potential evaporation) or less similar matches (insist on ten matches and accept that a few will contain non-trivial amounts of rain or lower rates of potential evaporation).

The previous issue notwithstanding, KNN produces a natural binomial distribution in its outcomes, as each element of the similar set becomes, essentially, a "voter" in a classifying election. Consider the simple case where $k = 9$. The most straightforward classification emerges by selecting whichever of the binary classes characterizes five or more members of the similar set. Alternatively, a confidence interval can easily be constructed:

$$P(A) = \frac{a}{k} \pm \frac{\left(\frac{a}{k}\right)\left(1-\frac{a}{k}\right)}{\sqrt{k}} z \qquad \text{(Equation 2.8)}$$

where P(A) is the probability that the current event falls within class *A*, *a* is the number of elements in the k-most similar set from class *A*. The Gaussian variable *z* specifies the confidence interval's scope, i.e. $z = 1.96$ for a 95% confidence interval and $z = 2.58$ for a 99% confidence interval.

To avoid a scenario in which all *k* elements of the similar set possess the same classification (i.e., a degenerate confidence interval case), two dummy examples are added to each similar set, one of which always reads "dry" and one of which is perpetually labeled "wet." In this manner, no classification will possess an unrealistic, zero margin of uncertainty.

2.3.2.3 Boosted Perceptrons

The final algorithm to be considered, also generally effective given limited information, is adaboost. Adaboost is an example of ensemble learning, which is used to generate multiple hypotheses (perceptrons) regarding the underlying function being modeled and to combine each resulting prediction and relative likelihood of accuracy (Russel & Norvig, 2010). In fact, adaboost is a mechanism to improve the performance of any weak learning algorithm, but for the purposes of this discussion, this section will discuss the "boosting" of a simple perceptron.

Perceptrons represent perhaps the most basic classifier, the simple linear separator. Consider, again, a series of examples, each characterized by a vector of features as given by equation 2.4. Next, a weight vector, is constructed as follows:

$$W = (\omega_1, \omega_2, \omega_3, \dots, \omega_n) \quad \text{(Equation 2.9)}$$

Such that:

$$\sum_{i=1}^{n} x_i \omega_i > \theta \quad \text{(Equation 2.10)}$$

θ serves as a threshold with which to delineate the two classifications. That is, the dot product of the weight vector and the input variables yields a scalar value. Values greater than θ output one classification while those below θ output another.

In the case of a problem involving modeling of the physical environment, especially one in which assessments are likely to be made with imperfect or incomplete information, it will likely be impossible to construct a linear separator (such as the one in equation 2.10) with every training example classified correctly. For this reason, adaboost improves model error by increasing the significance of those examples classified incorrectly, and decreasing the significance of those points for which the training algorithm was correct. In each iteration, a new linear separator is developed to minimize the weighted sum of squared errors. If certain examples are repeatedly misclassified, their errors will magnify until eventually, the best separator becomes one which will classify those examples appropriately.

The general theory of boosting stipulates that for any weak learner (one that classifies at least as well as a random guess), with a sufficient number of iterations, perfect classification will occur over all training examples, although performance in validation will plateau considerably sooner (Russell & Norvig, 2010).

Though there are numerous numerical patterns through which misclassified examples can be magnified and correctly labeled examples can be diminished. Adaboost's technique, used in this work, is given below.

Consider a series of *m* examples:

$$(X_1, C_1), (X_2, C_2) \ldots, (X_m, C_m) \qquad \text{(Equation 2.11)}$$

Where each X_i is a vector of features, akin to equation 2.4, and each C_i represents a binary classification. Initialize a weight for each example:

$$D_t(i) = \frac{1}{m}, \quad \forall_i, \quad t = 1 \qquad \text{(Equation 2.12)}$$

Determine a classifier, h_t, such that any vector of independent variables maps to -1 or 1. This is achieved via a threshold function as shown in equation 2.10. Stated mathematically:

$h_t(X_i) \rightarrow, \{-1,1\}, \forall_{i,t}$:

$$h_t = \underset{h_t \in H}{argmin}\, \epsilon_t \quad \text{where } \epsilon_t = \sum_{i=1}^{m} D_t(i) * \begin{Bmatrix} 0 \; if \; C_i = h_t(X_i) \\ 1 \; if \; C_i \neq h_t(X_i) \end{Bmatrix} \qquad \text{(Equation 2.13)}$$

Stop if $\epsilon_t \geq 0.5$, as this violates the principle of superiority to a random guess.

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \quad \text{(Equation 2.14)}$$

At this point, it becomes evident why the error rate ϵ_t of classifier h_t must be strictly less than 0.5, otherwise equation 3.14 produces an undefined result. Next, $D_t(i)$ is updated:

$$D_{t+1}(i) = \frac{D_t(i)e^{-\alpha_t C_i h_t(x_i)}}{\sum_i D_t(i)e^{-\alpha_t C_i h_t(x_i)}} \quad \text{(Equation 2.15)}$$

The denominator is simply a normalization term to ensure that the weights of each example sum to unity. The final classifier behaves as a weighted aggregate of each iteration:

$$H(X) = sign[\sum_{t=1}^{T} \alpha_t h_t(X)] \quad \text{(Equation 2.16)}$$

Once again, like the KNN algorithm, this algorithm can be manipulated slightly such that each classification falls between 0 and 1. First, rather than classifying in a binary manner of -1 or 1, simply choose 0 or 1. Second, in equation 2.16, a removal of the $sign(\,)$ operator will yield a weighted average of 0s and 1s. Thus, once again, a binomial distribution can be modeled and a confidence interval can be developed as shown in equation 2.8 for KNN, simply replacing $\frac{a}{k}$ with $H(X)$.

2.3.2.4 Application of the Algorithms

For each of the three algorithms discussed, certain problem-specific adjustments are required to ensure the production of meaningful results. First, it is important to recognize that there are certain summer days for which obtaining a correct classification becomes trivial, simply

because no precipitation has been seen for extended periods of time. No computational algorithm is required to inform the land user of soil conditions after a sufficient duration of dry conditions. Consequently, so as to avoid artificially enhancing the algorithm's accuracy, all examples for this case study in which no rain has fallen within the three previous days are eliminated from the testing set. Three days exceeds the longest drying period observed in these tile-drained agricultural fields, although a longer period would likely be necessary in less well-drained areas. This approach ensures that the algorithm's accuracy is measured only on examples for which the nowcast is non-trivial.

Second, as alluded to in the previous section, a substantial validation set is held aside from the collected dataset that is not used to calibrate the machine learning algorithms. By testing the best model only on training examples from previously unused data, external validity is verified and the probability of that model functioning effectively for future assessments of soil conditions increases. It is worth noting that, as the growing season of 2010 represents the entirety of the available data for this particular analysis, the training set contains 34 of the 109 days for which data have been gathered, a figure which could be perceived as fairly low considering that many machine learning algorithms train on much larger proportions of the data. This occurred because the beginning of the growing season in the spring contains considerably more wetting and drying events than the summer and fall. Should the training/testing ratio approach the more common 80%/20%, the validation set might not contain a single substantial period of storms. Moreover, random assignments to training and testing sets cannot be made in this case, as subsequent training examples are, naturally, dependent on previous examples, as drying involves a continuously varying, physical system. Thus, the training data consists of data

taken on or before June 24, 2010, and testing data consists of the growing season thereafter (ending September 24, 2010).

Finally, precipitation data older than 48 hours are not included in the models as independent variables. While there are certainly numerous locations in which rainfall's influence on the height of the water table, or even the dryness of the topsoil, might endure for weeks or even months, the fields at the South Farms test site are all tile-drained and empirical examination of the auto-correlation in soil moisture reveals that only the first two days affect dryness assessments.

2.3.3 Implementation

The machine learning algorithms constructed for this analysis were implemented in R, version 2.11.1, for Windows Vista. The implementation of the classification tree algorithm and the visualization thereof was aided, in part, by the rpart library. The remaining two algorithms (KNN and boosted perceptrons) were developed from scratch in R. The input variables are the precipitation data over the past 24 hours, the precipitation data from 24-48 hours previously, and the potential evapotranspiration estimates from the previous 24 hours. Potential evaporation estimates, from the ICN's hourly sensor readings are computed with a sliding 24-hour window. Precipitation readings, from NEXRAD data, are computed using inverse-distance-weighted interpolations (Shepard, 1968) from a 5x5 grid of 1km by 1km precipitation radar values that surround the test site.

In terms of computational demands, an office laptop (these algorithms were run on an intel i5, 2.53 MHz) was sufficient for timely execution. For classification trees, the tree could be placed in memory in a matter of a few seconds with individual conditions evaluated almost

instantaneously. For the KNN algorithm, a historical database could be entered into memory and back-tested in under one minute and individual queries occurred in under one second. Finally, for the boosted perceptron algorithm (which is the most computationally expensive), the multiple iterations required to develop the various linear separators could generally be implemented in under ten minutes, with individual conditions tested almost instantly.

It is worth noting, however, that these three algorithms respond differently to increasing scales. While a classification tree will take longer to construct with more examples and more features to model, once a tree is entered into memory, any new example can be classified very quickly, even with a very large tree. Classification time for a single example is approximately O(log n) where n is the size of the historical database, and thus, grows very slowly. For KNN, entering the database into memory is simply O(n), but back-testing any example requires a sort of the data, which is, at minimum, O(n log n). Thus, for very large datasets, the computational expense could become substantial. Boosted perceptrons can become time consuming to construct as datasets become very large, especially as the number of linear separators becomes large. Fitting the weights of the linear separators (the coefficients associated with each independent variable) through stochastic gradient descent runs in O(nm) time where m is the number of features and n represents the number of training examples. However, once a boosted separator is placed in memory, new examples can be classified almost instantly, as that computational time grows as O(m) where m is the number of features. Thus, so long as our variable space is unchanged, more data will not slow classifications.

2.4 Results

In this section, results will be presented for each of the algorithms investigated in this research, beginning with classification trees, proceeding to k-nearest-neighbors, and then finishing with the boosted perceptron results. The section will conclude with a comparison of the relative accuracies of each algorithm.

2.4.1 Classification Trees

For each algorithm, calibration and testing began as data arrived throughout the growing season. In the case of classification trees, during the very first attempts at constructing predictive models, the ability of classification trees to ignore irrelevant information in specific situations allowed for simple trees which outperformed other algorithms. However, as the data set became increasingly rich, the performance of the trees was rapidly surpassed by the remaining two algorithms. The best performing classification tree was trained on the 34 dates on or before 6/24/10 and tested on 75 days thereafter. This tree, which appears in figure 2.6, includes three features: the last 24 hours of precipitation, with more recent rain weighted more heavily (see equation 2.17); precipitation from the last 48 hours; and the most recent 24 hours of potential evapotranspiration.

$$Precip\_1 = \frac{\sum_{i=1}^{24} P_i * \left(\frac{i}{24}\right)^3}{\sum_{i}^{24} \left(\frac{i}{24}\right)^3} \qquad \text{(Equation 2.17)}$$

Where P_i represents the precipitation falling in hour i. In this weighting, $i = 1$ refers to rainfall from 23 to 24 hours before the time in question and $i = 24$ refers to the most recent hour's rain.

It is worth noting that this tree contains only five terminal leaf nodes, as the training data consist of only thirty-four examples. To avoid over-fitting the data and subsequent drop-offs in performance during validation, splitting was prohibited if a given node contained fewer than eight elements and no split was considered acceptable if fewer than three elements remained in any terminal leaf.

Note that in figure 2.6, if the inequality is *true*, the algorithm proceeds to the left branch of each bifurcation. For example, if "Precip_2," which refers to all precipitation falling within the previous 48 hours (mm), is too low (not much rain recently), the algorithm chooses "DRY" and does not even consider any other features.

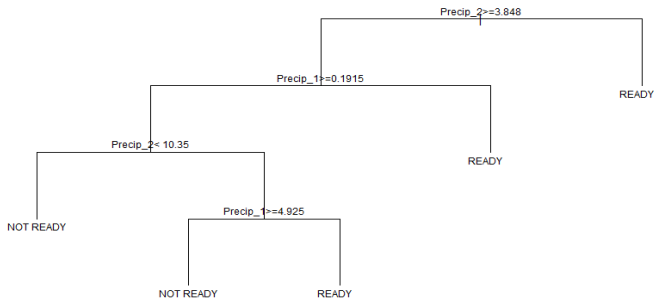


Figure 2.6 – The "best" classification tree

This tree's performance (see figure 2.7), at roughly 88% on the validation data falls below that of other algorithms. It is important to recognize that the actual dryness is scaled 1-5, while the algorithm attempting to classify soil conditions only returns two values ("dry" or "wet"). A "correct" classification occurs if and only if both the blue and red lines fall within the same shaded region. Thus, for example, after one error in early September, the algorithm is correct on every subsequent day, despite the distance between the red and blue lines. The one

day on which the blue and red lines overlap (9/22) is no more "correct" than the day before or after.

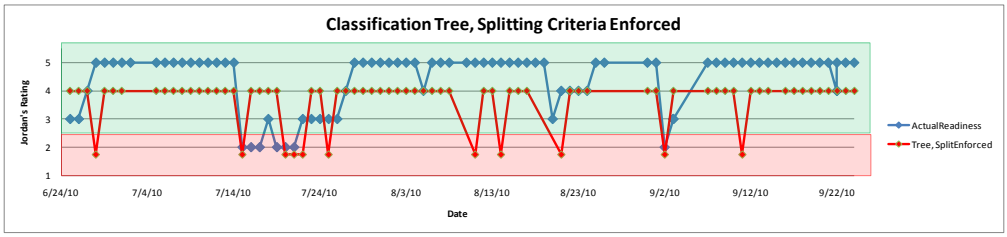


Figure 2.7 – The "best" classification tree, splitting criteria enforced, actual vs. predicted

Despite the 88% accuracy of the tree in figure 2.6, careful observation reveals that potential evaporation data never becomes a splitting criterion. Evidently, there is always a more effective means of increasing information gain, considering that classification trees must consider variables one at a time. With this in mind, a "complete" tree can be constructed, splitting until either: (a) nodes contain a single example, or (b) no improvement is possible on training data. Given consistently labeled examples, a decision tree ought to correctly classify all training data. However, over-fitting a tree in this manner can lead to deteriorating performance in validation.

Once again, decision trees possess the inherent advantage of focusing upon the variable which produces the maximum quantity of variance at any given stage of the decision-process, yet, unfortunately, they fail to utilize all relevant features without over-fitting the data. While the aforementioned strength outweighed the weakness during initial testing (when data limitations were extreme), the following algorithm improved upon them.

2.4.2 K-Nearest-Neighbors

Next, the best binary classifier is constructed using the k-nearest neighbor algorithm with, in this case, three input variables. As described for the classification tree algorithm, the first variable represents an aggregation of all precipitation within the most recent 24-hour period, weighting more recent rainfall more heavily. The second represents a simple aggregate of all rainfall within the most recent 48-hour period. The third is the potential evapotranspiration over the most recent 24 hours. For the sake of visual clarity, in figure 2.8 the red line for KNN prediction is scaled closer to five when increasing numbers of the similar set classify the example in question as "dry" and closer to one when more members of the similar set classify to "wet." The scale system used ensures that 5.5/11 votes would classify to 2.5 – the boundary between the green and red shaded regions. In other words, a 50/50 split of the similar set would yield a forecast on the threshold between dry and wet. In this case, only five errors in classification occur, bringing our accuracy up to 93% within the validation set. Moreover, the cases in which errors occur are those for which the eleven nearest neighbors have split votes.

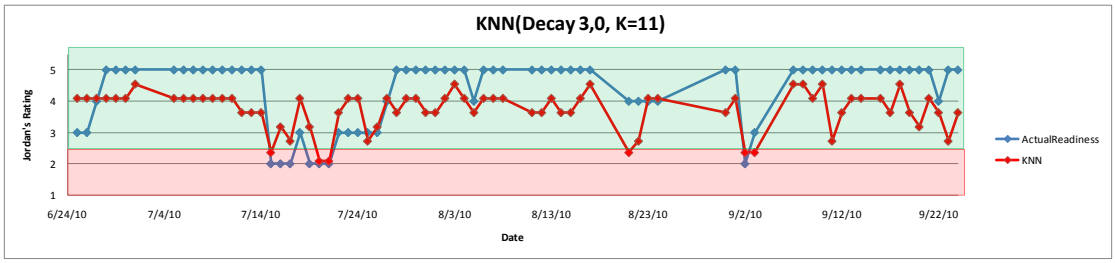


Figure 2.8 – KNN, actual vs. predicted

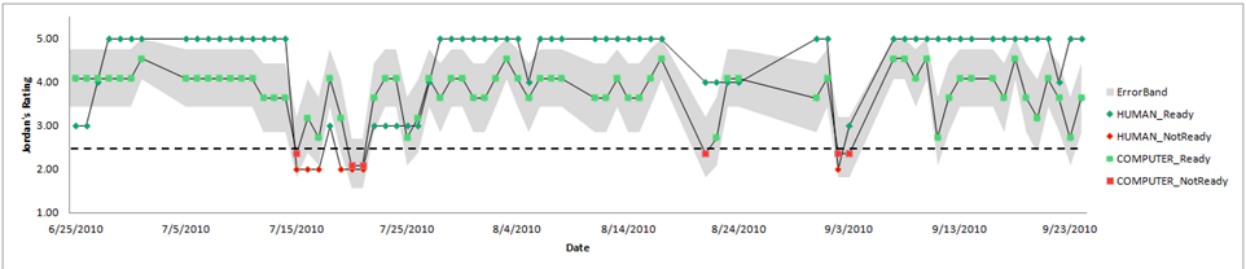


Figure 2.9 – KNN, actual vs. predicted, with binomial confidence interval (gray band)

In Figure 4.4, a binomial, 99% confidence interval is computed using the KNN votes and dummy votes as shown in equation 3.8. Note that the five errors (cases in which the human's assessment is green and the computer's prediction is red or vice-versa) all occur where the boundary (dotted-line) between dry and wet falls within the confidence interval. That is, the algorithm never errs outside of its designed margin for error in seventy-four classifications. Real-time use of such an algorithm could occasionally cause uncertainty in cases where a field is truly "dry" by predicting a probability of that classification near 50%. However, if the algorithm predicts that a field is viable at a given point in time with a high probability, a farmer could act with confidence that his/her site is, in fact, dry.

2.4.3 Boosted Perceptrons

Lastly, binary classifications are developed using the boosted perceptron algorithm with the same three inputs as deployed for the k-nearest-neighbor algorithm. Like KNN, the algorithm's ultimate prediction is the aggregate of multiple "votes," in this case, individual perceptrons with weights as shown and calculated in equation 2.14. Thus, any classification is scaled between 0 and 1 and can be visualized on the 1-5 scale such that any classification greater than 0.5 (dry) appears above 2.5 on the y-axis, as shown in figure 2.10.

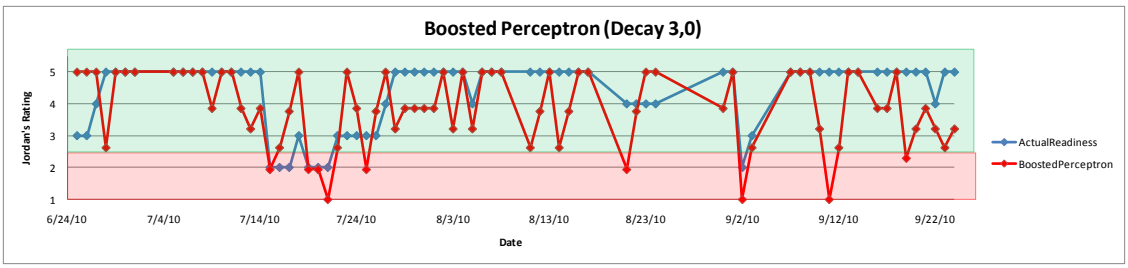


Figure 2.10 – Boosted Perceptron, actual vs. predicted

The performance is quite similar to the KNN algorithm, as 92% of all points are classified correctly, and several dates with errors overlap between KNN and boosted perceptrons. However, the KNN forecasts generally yield estimates closer to 50% while the boosted perceptron algorithm outputs more extreme predictions (closer to 0 or 100%) – this is a mathematical relic of the fact that the most similar examples in KNN often contain both "dry" and "wet" days, as well as off-setting dummy votes. The boosted perceptrons, without the advantage of a "dummy" vote, often will classify 100% toward one designation or the other. Encouragingly, in 34/35 (97%) of the cases in which the boosted perceptron algorithm is "certain," that is to say that each individual classifier agrees on the classification, the ultimate conclusion is correct. The erroneous case will be discussed in a later section.

As was the case with KNN, a confidence interval can be constructed around the 0 to 1 classification found with boosted perceptrons, then scaled from 1 to 5 for visualization purposes. The results are shown in figure 2.11.

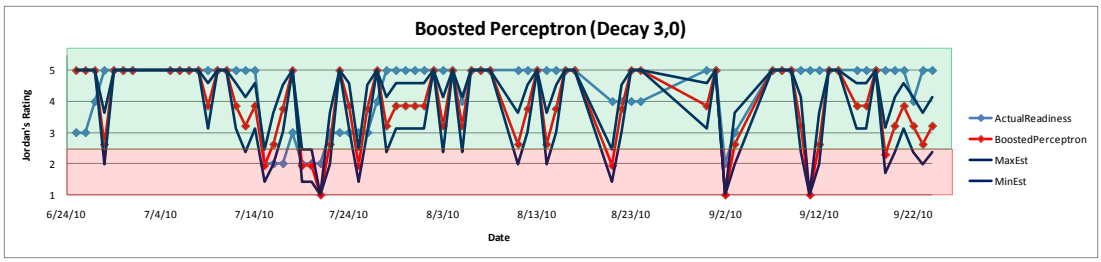
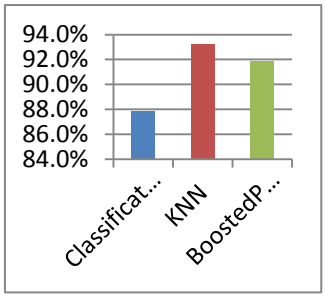


Figure 2.11 – Boosted perceptron, actual vs. predicted, with binomial confidence interval

As is evident from the above chart, on several dates the confidence interval is degenerate, that is to say, of nil width. Despite this, there are only two cases in which a misclassification occurs without the confidence interval crossing the boundary between dry (green shaded region) and wet (red shaded region).

2.4.4 Algorithmic Comparison

Finally, in Figure 4.7 the accuracy of the three algorithms is compared. The three bars compare the accuracy of each of the algorithms tested on validation data that was not used for training. The classification trees have the worst performance of the three algorithms, misclassifying several examples that the other two algorithms classify correctly, while KNN performs the best – with even stronger performance after a confidence interval is introduced.



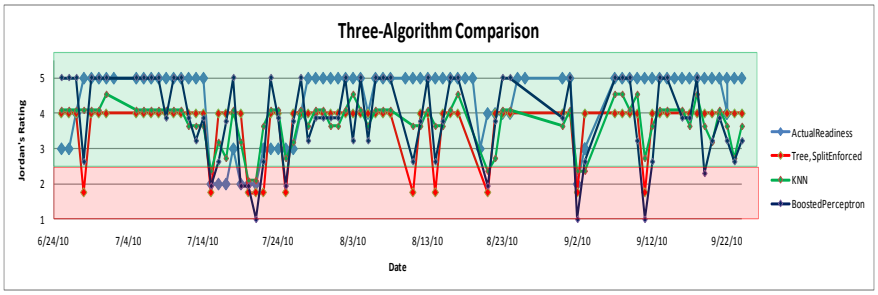


Figure 2.12 (left), Accuracy of the three algorithms
Figure 2.13 (right) – Comparison of predictions from the three algorithms

To explore these differences further, Figure 2.13 compares daily predictions from the three algorithms with the farmer's assessments. It appears that there are a handful of "aberrant" examples within the training set (8/21, 9/2, and 9/11), in which the actual assessment is "dry" despite higher levels of precipitation and lower levels of potential evaporation – perhaps the qualitative assessments are inconsistent in a few cases. Given the limited number of training examples, these handful of unusual examples can play a much more substantial role in the ultimate classification of a validation example.

KNN narrowly outperforms the boosted perceptron algorithm, failing to err even once outside of the confidence intervals constructed (Figure 2.12). However, the boosted perception is much more likely to classify strongly, i.e. return a probability closer to 0% or 100%. This is because with KNN, where k=11, given the limited training data, it is a virtual certainty that at least a couple of those similar examples deviate from the rest. This ensures that at most 9 or 10 examples will classify the same way, even before adding dummy votes. However, the boosted perceptron, by construction, develops hypotheses which, at times, return classifications at the extremes (and is 97% correct when it does). Future growing seasons with additional data at more sites will determine whether KNN or boosted perceptrons emerges as the best practice for predictions of this nature.

Ultimately, of the three algorithms examined, two represent viable and effective means of modeling the dryness of a specific site given user-defined assessments – KNN and boosted perceptrons. Both fall between 91 and 94% accuracy, even after removing those days for which classification would be easiest due to a lack of recent precipitation. Furthermore, the errors made by these two algorithms tend to align, illustrating that other factors may explain the erroneous results beyond simple model inadequacy. On three of the six days (7/17/10, 7/25/10,

9/19/10) in which the boosted perceptron algorithm erred, no time stamp was given for the qualitative measurement. One of these days (7/17/10) represented one of the five misclassifications for KNN as well. As a result, the algorithm selected a default time (9:00 AM) upon which to base predictions. As these measurements were varied in terms of their time of day from early morning to late afternoon, it is wholly plausible that the algorithm was correct at the time of day assumed, but ultimately incorrect given the unknown time at which the measurement was actually obtained.

Furthermore, several errors also occurred when the field conditions themselves were borderline cases between two classifications. In this case, minor qualitative inconsistencies in human assessment will lead to incorrect classifications.

Finally, the qualitative assessments lack any internal means of verification – as a result, it is possible that a data point is erroneously reported. In the case of 9/11/2010, substantial precipitation fell within the most recent 24-hour period as well as from 24-48 hours. Additionally, rates of potential evapotranspiration were below half of the growing season average over the 24 hours preceding measurement. The classification tree and boosted perceptron classified the field as "wet" while KNN chose "dry." The qualitative conditions were reported as a "5" – i.e. a fully dried soil. This value is dubious, yet unverifiable retroactively, and resulted in errors for two algorithms (as well as the one case in which the boosted perceptron predicted a 0% or 100% probability of dryness and was incorrect).

While boosted perceptrons currently do not match the performance of KNN, the level of accuracy is only marginally inferior. By using stochastic gradient descent (Russel & Norvig, 2010) to fit the parameters of each individual hypothesis, dimensionality issues create less of a concern. Of course, with each additional variable, the opportunity to over fit increases as well.

2.5 Discussion & Conclusions

Despite occasional errors, which are uncommon and seem well explained by issues with data gathering, qualitative assessments, and missing information, two algorithms perform at 92% and 93% accuracy on validation data, and exceed even those figures when a confidence interval is introduced. It is possible that these statistics would improve as each successive growing season provides an increasingly rich data set, incorporating a greater variety of conditions and locations. Each algorithm needs further validation, but these results suggest that nowcasts with high levels of accuracy are likely attainable.

What is truly encouraging about these results is not only their accuracy, but the limited information requirements. Many agricultural plots in the Midwest and elsewhere are characterized by relatively efficient drainage from tile drains. In sites such as the test location in Urbana IL, USA, knowledge of rainfall over the most recent 48 hours is wholly sufficient to model growing season wetting conditions. The API approach requires precipitation histories in excess of two weeks. The diagnostic soil moisture equation requires an indeterminate precipitation history that can be two to three months long. While this may ultimately provide a more precise estimate of soil moisture, in terms of tangible decision-support, the method presented in this chapter offers sufficient performance at lower data costs. Moreover, both the diagnostic soil moisture equation and API approaches (and hydrologic models) pose difficulties in addressing precipitation that arrives as snow, settles in snow packs (not wetting the soil), then melts at a later date. When the precipitation history demanded is quite lengthy, a prediction during the planting season (e.g., early April at our test site) is likely to contain days of precipitation that have arrived as snow/ice and will disrupt results.

Furthermore, lengthier predictive windows may cause increasingly persistent errors. Consider a model that employs a 50-day precipitation history. If, on day X, the model is incorrect, on day $X + 1$, the model is likely to be incorrect as well, as after all, 49 of the 50 days of precipitation data used for day X are still employed for day $X + 1$. For this reason, an erroneous data point, a phenomenon whose impact is poorly addressed by the model, or simply a slight miscalibration, can endure for extended periods. If only two days of history are needed, a error on day X is likely to be resolved on day $X + 1$, and is entirely irrelevant by day $X + 2$.

Given the construction of these algorithms using free software (R) and computational requirements that are feasible on a conventional laptop, subsequent seasons and locations of data could be integrated without the need for structural changes to the software. Many hydrologic models for such wetting/drying analysis require frequent recalibration to avoid cumulative errors (Jones, 2004). However, by using individuals' personal standards of "wet" and "dry," these machine learning tools are naturally, and constantly recalibrating, adjusting to the judgment of the decision-maker. By asking decision-makers to provide feedback on the accuracy of the predictions (e.g. through a mobile application), each input augments the set of training data from which the algorithm can base its decisions. Errors do not propagate, but rather, are smoothed by increased knowledge.

Additionally, since these models are developed through machine learning protocols rather than physical soil mechanics, provided that a repository of consistently labeled training examples exist, such algorithms can be constructed for soil drying in many agricultural contexts. As the results indicate, public data yield relatively accurate predictions, implying that these algorithms can be constructed without the cost and logistical complexity of adding remotely-accessed sensory grids. Both the diagnostic soil moisture equation, API, and hydrologic approaches

require soil moisture sensors to initially calibrate the model. In the case of machine learning models of qualitative assessments, the individual decision-maker is sufficient as a calibrating tool – and if their opinion of what "dry enough" entails changes, so too can the model.

Machine learning algorithms, being sufficiently flexible, should be able to adapt their parameters to any established, consistent, dryness metric. Such predictions could be worth investigating in a wide variety of circumstances (e.g., agricultural land use, confined animal feeding operations, and natural resource management). If constructed solely with public data, the number of site users who could conceivably benefit is substantial.

This work, despite its broad applicability, focused upon a single location – a topographically flat, tile-drained field, with constant soil types, in a Midwestern climate. Certainly, those characteristics are not shared by all locations in the United States in which such decision-support tools might prove useful. To this end, it becomes necessary to develop a classification system which can cluster catchments into groups of similar climate properties – allowing model construction to be location-appropriate. The next chapter will address the development and application of such a system to over 400 catchments within the continental U.S.

CHAPTER 3

NATIONAL-SCALE CLASSIFICATION

3.1 Introduction: Relationships Between Regime Curves and Flow Duration Curves

This chapter fits within a broader research initiative amongst several researchers to better understand the physical drivers of observed regional patterns of flow duration curves. One paper from this initiative, by Cheng et al (2012), focuses directly on the FDC and approaches the problem empirically, fitting the parameters of a gamma distribution to describe its shape. Another, from Ye et al (2012), adopts a top-down modeling approach to explore the dominant process controls upon the regime curve and their subsequent relationship to the FDC. The final paper of this initiative, (Yaeger et al, 2012), synthesizes the insights from the different perspectives of Cheng et al (2012), Ye et al (2012), and the work discussed within this chapter.

The chapter begins with a discussion of hydrologic similarity, focused on similarity of regime behavior, which encapsulates the mean within-year variability of key hydroclimatic variables. In this respect the four key indices that will be used to quantify hydrologic similarity are then presented and justified. This section is followed by a detailed presentation of the methodology used to construct the decision tree, including the similarity metrics used and a demonstration of the Iterative Dichotomiser (ID3) clustering algorithm. The results of the implementation of the decision tree will be presented next, discussing each of four quadrants of the tree in detail, and focusing upon six pivotal clusters that happen to describe the majority of the catchments analyzed. Having established clusters of similar catchments, the robustness of the classification tree is verified by reconstructing the decision tree using a 200-catchment subset of the 400 catchment database used to develop the catchment classification. The chapter

concludes with a hydrologic assessment of the catchment classification achieved, including lessons learned and questions left for future work.

3.2 Similarity: What Does it Mean Conceptually?

Since the focus of this paper is on the catchment's regime behavior, two catchments will be considered hydrologically similar if their regime behavior can be deemed similar. In this paper, four key similarity indices will be used to characterize the similarity of regime behavior, and are defined and discussed in detail later in this section. These include (i) aridity index, a measure of aridity that, to first order, determines the annual water balance, (ii) a seasonality index that quantifies the strength of seasonal variability of precipitation within the year, (iii) The timing (mean date) of precipitation peak within the year, and (iv) the timing (mean date) of runoff peak within the year. Since the climate of the continental United States is such that the seasonal variation of energy (and temperature) is relatively uniform across the country, the timing of precipitation is effectively a measure of the phase difference between the seasonality of precipitation and potential evaporation. On the other hand, the timing of the runoff peak (especially in relation to precipitation and potential evaporation) captures the mechanisms of storage (in soil water or snow storage) and release (in terms of subsurface drainage or snowmelt). In this sense the similarity indices provide a first order mapping towards the regional variations in dominant processes highlighted in the parallel study of Ye et al. (2012).

3.2.1 An Example of Regime Behavior

Figure 3.1 presents the daily regime curves of precipitation, potential evaporation, total runoff for a fairly typical Midwestern-American catchment, located in Kansas.

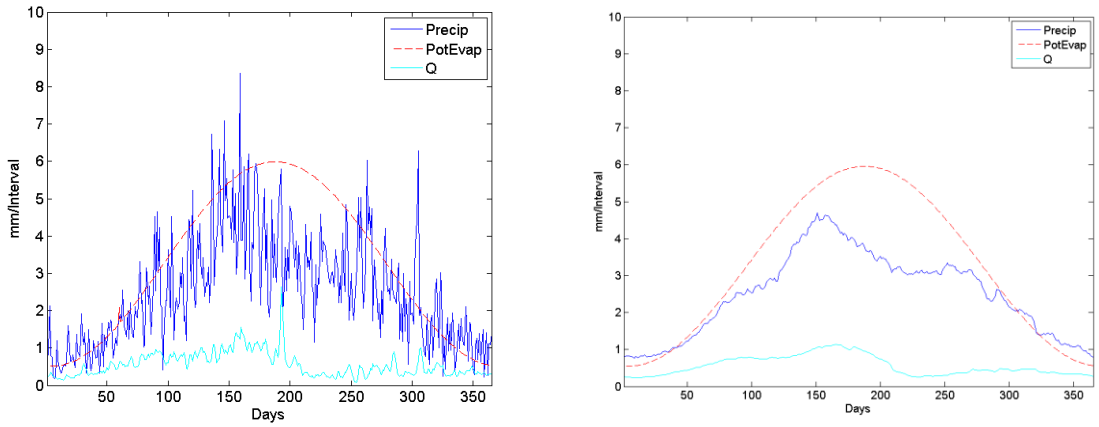


Figure 3.1: (a) Daily Regime Curve, and (b) 30-day Moving Average, Marais Des Cygnes River, Near Ottawa, Kansas, USA.

The left image (Figure 3.1a) is obtained, using MOPEX daily data from 1948 to 2001 (Sivapalan et al., 2011; Cheng et al., 2011)[1], using ensemble averaging by calendar day. While Figure 3.1(a) does provide useful information about the within-year (daily) variability of the chosen variables, for the purpose of catchment classification in this paper, a sliding, 30-day moving average is generated, as shown in Figure 3.1(b). Equation 3.1 captures this smoothing,

$$P_i = \frac{\sum_{i-15}^{i+15} P_i}{\sum_{i-15}^{i+15}(1)} \qquad \text{(Equation 3.1)}$$

where P_i represents the average precipitation for day *i* of the year; as a point of clarification, this calculation is circular. Many hydrological analyses (Koppen-Geiger, Haines et al, and others) deploy monthly regime data to depict seasonal patterns of rainfall and runoff. A 30-day moving average achieves this idea of a 30-day window without creating arbitrary monthly boundaries. In

[1] MOPEX data obtained from http://voda.hwr.arizona.edu/mopex/

classifying catchments on the basis of the daily regime curves of climatic and runoff data, in this paper we will focus upon images like this one (Fig. 3.1b) for all 428 catchments within the MOPEX database (Duan et al., 1992). The proposed classification scheme will be built around four key indices, each extracted from the smoothed regime curves of the type presented in Fig. 3.1b.

3.2.2 Similarity Indices Used

In the spirit of Koppen-Geiger, a key objective of this research is the classification of regime behavior using an absolute minimum quantity of data, on the basis of four very simple and widely available similarity indices. To estimate these four indicators, three daily time series are required: precipitation, potential evaporation, and total runoff. The first index is the aridity index (*Ep/P*), the ratio of annual potential evaporation to annual precipitation; it measures the competition between energy available and water available, and is seen as a good first-order indicator of runoff ratio (ratio of annual runoff to annual precipitation). Note that the phase of *Ep/P* is not addressed because within the continental U.S., every catchment's *Ep* curve peaks within a couple of weeks during the summer, and *Ep* is very low during winter months; thus the curve's amplitude is subsumed by the value *Ep/P*. The seasonality index and maximum day of precipitation are both estimated from the daily precipitation time series. The seasonality index measures the strength of within-year (seasonal) variability of precipitation, and is zero if the precipitation is uniform throughout the year. The timing of rainfall peaks is a reflection of the phase difference between the timing of the precipitation peak and that of potential evaporation (Milly, 1994), given that in the continental United States, the timing of potential evaporation's peak is uniform spatially. Finally, the timing of maximum runoff accounts for the response of

the catchment to the interactions between precipitation and potential evaporation. The timing of runoff allows for differences in storage and release processes between different regions, owing to distinctions in topography, snowfall, snow storage and melt processes, and also differences in the physiological responses of vegetation. The decisions with respect to the four indices are justified in terms of understanding the interplay between wetting and drying, and the timing separating rainfall from runoff, as discussed in Cheng et al and Ye et al (2012). Within the United States, any three indices are insufficient to understand the nuanced behavior of the catchments we examined, but the addition of a fourth (at least for the vast majority of MOPEX catchments) resolves the discrepancies.

In essence, the four indices represent answers to the following four questions:

- "Is this catchment very humid, somewhat humid, temperate, somewhat arid, or very arid?"
- "Is rainfall relatively consistent year-round, somewhat seasonally dependent, or highly seasonally dependent?"
- "When, during the year, is rainfall greatest?"
- "When, during the year, is streamflow greatest?"

Other variables, such as runoff ratio (*Q/P*) were considered, but ultimately not adopted because they were correlated with other variables (*Ep/P*) and failed to improve the quality of classification. Our classification system was reconstructed after the omission of each of the four features to verify that, in fact, all four features are necessary. Further justification of the four features selected is available in Appendix A.

With respect to the independence of the four features, seasonality and aridity index are almost entirely independent (r^2~0.14). Seasonality and date of max precipitation are fully independent ($r^2 < 0.01$). Seasonality and date of max runoff are almost entirely independent (r^2~

0.14). Aridity index and date of max precipitation are independent (r^2~ 0.05). The same is true for aridity index and date of max runoff (r^2~0.06). Though one might suspect the date of peak precipitation and peak runoff to be related, the r^2 value connecting the date of maximum precipitation and the date of maximum runoff is only 0.21. Though there are clusters where the maximum runoff follows the maximum rainfall by a few days or weeks, there are also numerous catchments with virtually constant annual rainfall, yet still characterized by a defined runoff peak. Finally, there are catchments that receive their highest rates of precipitation during fall/winter, then store that water in snowpacks, yielding peak runoff in April, May, or June.

While there are other features that are relevant in understanding the behavior of a given catchment (proportion of precipitation as snow, etc), these concepts are included, at least in large part, in the four features chosen. While these four features are sufficient for our purposes, future research may consider adding further indicators to improve specification within certain regions.

3.2.2.1 Aridity Index: Dry or Wet?

Figure 3.1 shows that in this catchment in Kansas, the daily potential evaporation rate is almost uniformly in excess of the daily precipitation rate throughout the year. The aridity index (E_p/P) is estimated by summing the mean daily rates of potential evaporation (PE) over the 365-day time series and dividing it by the sum of the mean daily precipitation rates over the same 365-day period. The aridity index thus measures the competition between energy available and water available annually: $E_p/P > 1$ for arid (dry) catchments whereas $E_p/P < 1$ for humid (wet) catchments.

Figure 3.2 presents regional patterns of the aridity index for 428 catchments belonging to the MOPEX database. It shows that eastern catchments tend to be largely humid (except in the south), whereas Midwestern catchments to be mostly semi-arid, becoming more arid as they approach the Rocky Mountains and the desert south-west, and to become humid again in the Pacific Northwest. Essentially one finds systematic east-west (and north-south) trends in the aridity index, compounded by outliers in the south-east and north-west.

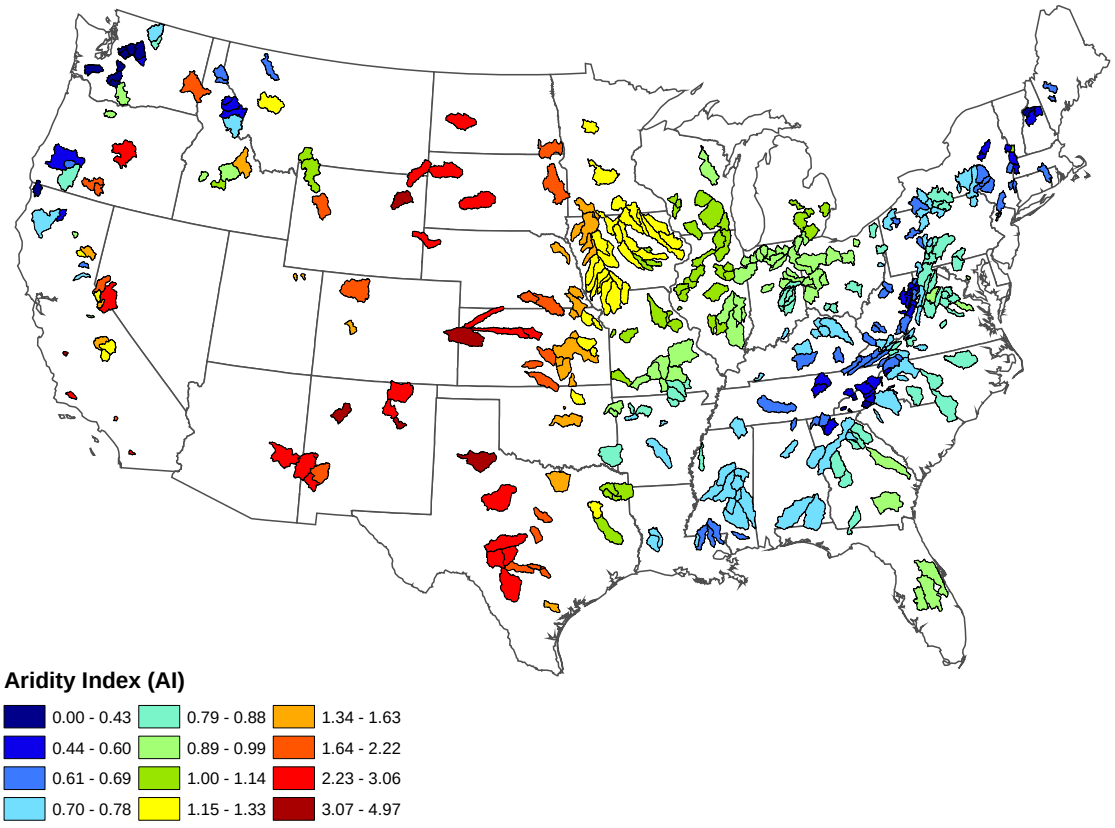


Figure 3.2 – Spatial Distribution of the Aridity Index (E_p/P)

3.2.2.2 Seasonality Index: Is Precipitation Uniform or Periodic?

Figure 3.1 presented a catchment whose precipitation and also runoff response both exhibited significant seasonality, with rainfall being much higher during the summer than during

winter months. This is a feature exhibited by a significant number of catchments belonging to the MOPEX database. Potential evaporation is also highly seasonal, although in this case there is very little phase difference between precipitation and potential evaporation. The relative magnitudes of precipitation and potential evaporation are likely to have an impact on runoff regime, and must be accounted for in the classification scheme. Walsh and Lawler (1981) defined a seasonality index for precipitation on the basis of average monthly rainfall values, which in essence a measure of within-year variance. In this paper we use an adaptation of Walsh and Lawler to accommodate the 365-day smoothed precipitation regime curve as follows:

$$\frac{\sum_{i=1}^{365}\left|P_i-\frac{\sum_{i=1}^{365}P_i}{365}\right|}{\sum_{i=1}^{365}P_i} \qquad \text{(Equation 3.2)}$$

where P_i represents the value obtained from Equation 3.1. The seasonality index helps to distinguish those regions in which precipitation is highly variable seasonally from those in which rainfall is comparatively uniform throughout the year.

Figure 3.3 presents the spatial distribution of the estimated seasonality index across the USA. In the eastern part of the country, precipitation is fairly uniform year-round with the exception of three catchments located in southern Florida. Moving westward, the seasonality index tends to increase, becoming moderately seasonal in the Midwest (mid-continent) and peaking in those catchments near the Pacific. While there are a few catchments that do not follow this trend in the northern Rocky Mountains, this general trend remains consistent.

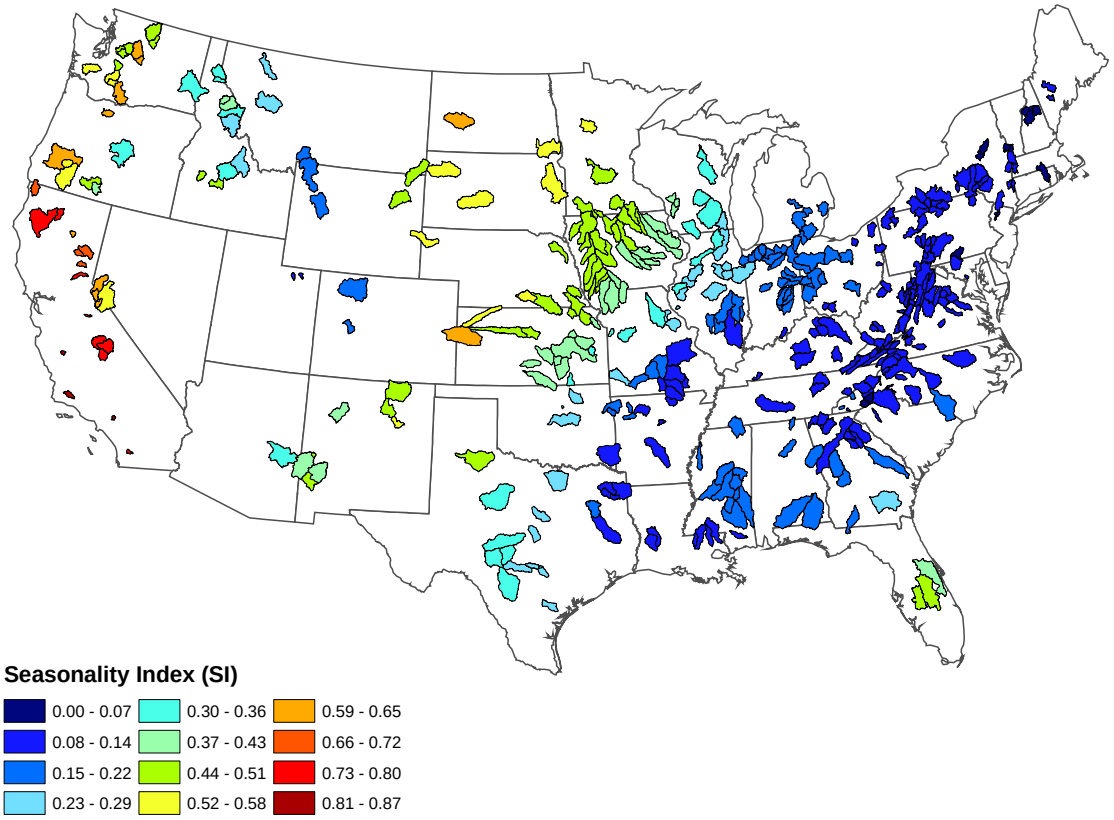


Figure 3.3 – Spatial Distribution of the Seasonality Index

3.2.2.3 Day of Peak Precipitation: In-Phase or Out-of-Phase with respect to PE?

With the seasonality index defining the strength of seasonal variability of precipitation, another key feature is the timing of the maximum precipitation within the year. In this case, the metric we use is the date (from 1 to 365) on which the smoothed precipitation regime curve has its peak. Given that the timing of the peak of potential evaporation is uniform throughout continental United States, the timing of the precipitation peak serves to focus attention to the phase difference between these climate variables, i.e., whether precipitation seasonality is in phase with that of potential evaporation (e.g. precipitation peaks during June or July), is out-of-

phase (precipitation peaks during December or January), leading PE somewhat (precipitation peaks during Spring months) or lagging PE somewhat (precipitation peaks during Fall months). As an aside, it is important that this similarity index be estimated from a regime curve obtained with a suitably long moving window to avoid mischaracterizing a catchment. These distinctions are important, since the phase differences between the seasonality of water input and energy input impact storage and release mechanisms, and can thus impact the magnitude and timing of runoff as well.

Figure 3.4 presents the distribution of the day of maximum precipitation using a circular color coding, i.e., if the day of maximum precipitation happens to fall on December 31 for a catchment, then the similarity index is quite similar to another catchment with its precipitation peak falling on January 2, even though the timing index will be "365" for the first catchment and "2" for the second catchment. Although numerical values are different, they are actually similar with respect to the timing of the precipitation peak. The color coding reflects this similarity.

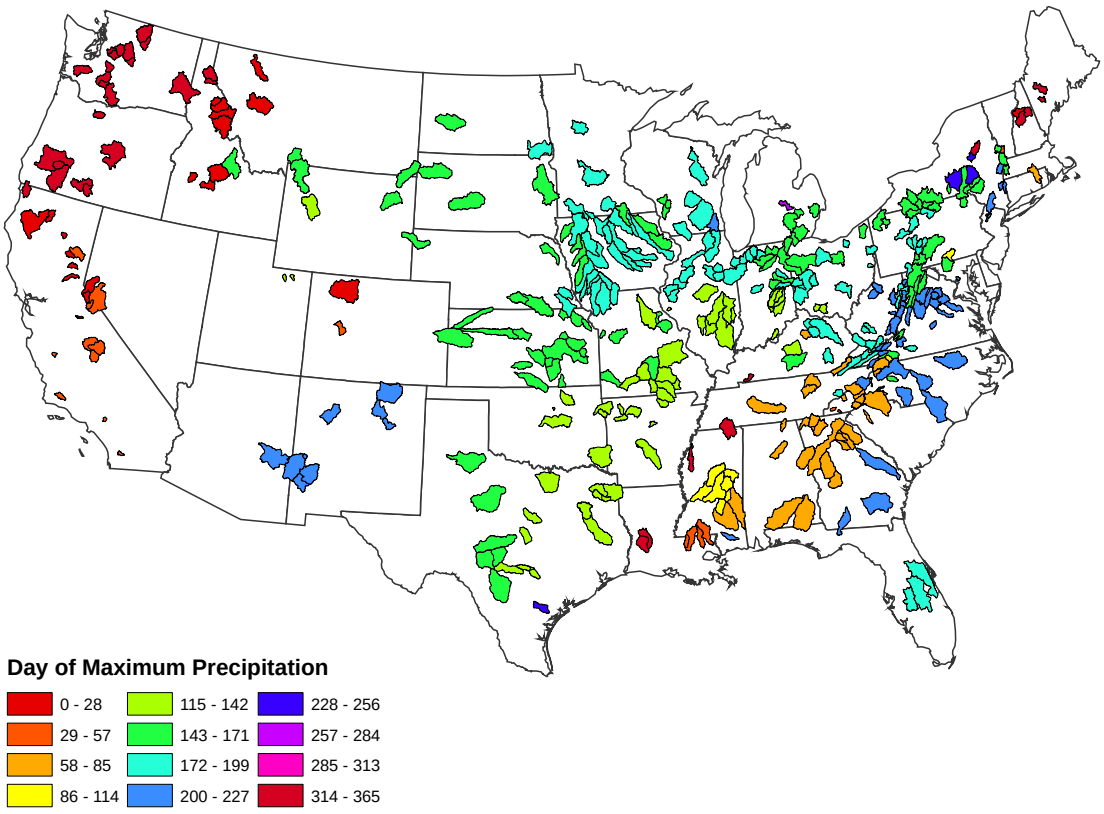


Figure 3.4 – Spatial Patterns of the Day of Peak Precipitation Day (1-365)

In this case, the east-to-west trends seen in the case of aridity index and seasonality index no longer hold. Although the Midwestern regions see precipitation peak in the late-Spring to early-Summer, there is much more variability across the continent, creating smaller clusters that are less-defined by longitude and latitude alone. The southern Appalachians are quite different from their northern, snowy counterparts, the Pacific coast displays a notable gradient north to south, and several catchments in the monsoon-influenced southwest distinguish themselves from their snowmelt-driven neighbors to the north, and their hurricane-influenced neighbors to the east.

3.2.2.4 Day of Peak Runoff: Role of Catchment Storage and Release Processes

Analogous to the day of peak of precipitation, the day of peak runoff (1-365) is the final piece to the classification puzzle. In this case, the differences between the catchments reflect not only the magnitude and phase differences between precipitation and potential evaporation, but also the transformations that happen at the land surface (storage and release processes, including below-ground soil and groundwater processes and above-ground snow storage and snowmelt), as illustrated by Milly (1994).

The results are presented in Figure 3.5, once again using a color coding scheme that is circular (1-365). As with the day of peak precipitation, we observe clusters that are not solely longitude- or latitude-driven, including considerable local variations that may reflect landscape heterogeneity. The Pacific Northwest distinguishes itself due to the out-of-phase relationship between precipitation (peaks during winter) and potential evaporation (peaks during summer). In the Midwest and along the east coast there is considerable heterogeneity, and in some cases even adjacent catchments show differences in runoff timing. Along the Appalachian mountains in the eastern half of the continent, runoff peaks appear in early spring, presumably driven by melting snow and spring rainfall.

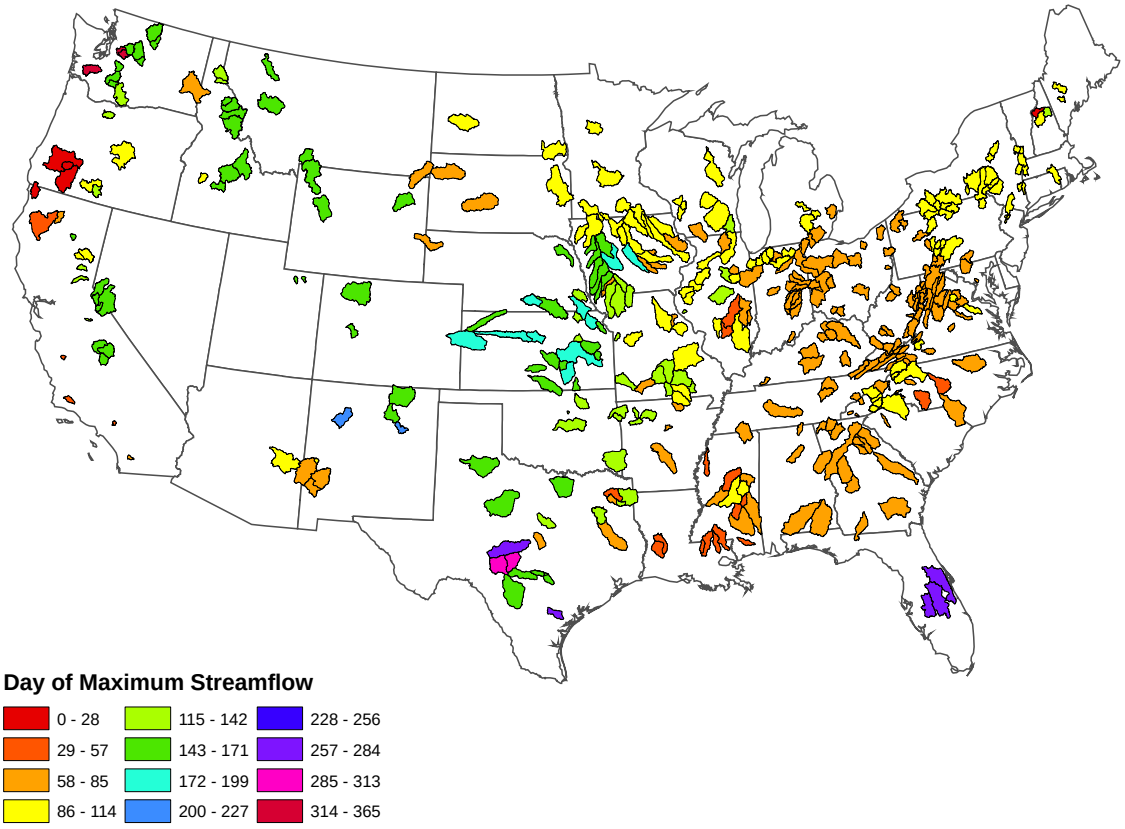


Figure 3.5 – Day of Peak Runoff (1-365)

3.3 Developing a Catchment Classification System

3.3.1 Decision Trees for Grouping Catchments

The goal of this section is to describe the methodology adopted in this paper to "group" catchments exhibiting similar *regime behavior*, and separate them from those that are different. Figs. 3.2 through 3.5 also exhibited certain regional trends across the continental USA with respect to each of the 4 similarity indices we had considered, and some level of clustering. In the same way, if the regime curves of the type presented in Fig. 3.1, generated for each of the 428 catchments in the MOPEX database, are superimposed upon a large map of the USA, one could

see regional trends, including the emergence of distinct clusters of similar regime behaviors (at least qualitatively). Is there a connection or possible mapping between the former and the latter? Our hypothesis in this paper is that a combination of the 4 similarity indices governs the regime behavior and can be the basis of their classification.

Considering that ultimately we want to develop a catchment classification system on the basis of regime behavior, and the fact that we have 4 different similarity indices that might collectively determine similarity of regime behavior, how can we develop a robust classification system? One way to develop such a classification system is via "decision trees" that can recursively divide the 428 catchments into self-similar groups in such a way that, at each step in the decision tree, the variability of a catchment attribute within each group is less than the variability between groups. The reason for a classification tree rather than another clustering algorithm, of which there are several in the literature (neural networks, nearest-neighbor algorithms, genetic algorithms, etc), was that this structure allows for qualitative insights to emerge along the way rather than a black box that delivers groups without explanation. With this method, as "observers" of the algorithm, we can see what splits occur on what values at what point in the process, allowing us to ask "What is the most important, most distinguishing characteristic for all U.S. catchments?", "What if we only consider the non-seasonal half?", etc. Constructing such trees requires a suitable metric – a mathematical definition of similarity (of regime behavior) that can be deployed for any group of catchments – a metric that encompasses the four key indices.

3.3.2 Metric of Regime Similarity

Each of the four similarity indices, the aridity index (A), seasonality index (S), day of peak precipitation (τ_p) and maximum runoff day (τ_q) show considerable variability across the catchments, which can be expressed in terms of a variance measure. For S this is straightforward, with the estimation of the standard deviation obtained from:

$$\sigma_S = \sqrt{\frac{\sum_{i=1}^{n}(S_i - \mu_S)^2}{n-1}} \quad \text{(Equation 3.3)}$$

where S_i is the seasonality index for catchment *i*, μ_S is the its mean over all catchments, and n = 428 is the number of catchments. An analogous estimate can be obtained for the standard deviation of the aridity index.

In contrast, for τ_p and τ_q, this estimation is not as straightforward. This is due to the circularity of the timing of the two peaks (i.e., 1-365), as in the case of four catchments whose values for τ_p are 361, 364, 359, and 3. To overcome this, we transform τ_p and τ_q into new variables C_1 and C_2 both of which naturally fall between -1 and +1, and overcome the circularity problem.

$$C_1 = sin\left(\frac{M_p}{365} * 2\pi\right), C_2 = cos\left(\frac{M_p}{365} * 2\pi\right) \quad \text{(Equation 3.4)}$$

By estimating σ_{C_1} and σ_{C_2}, the variances of C_1, and C_2, respectively, we can then estimate the standard deviation of τ_p, σ_{τ_P}, as follows:

$$\sigma_{M_P} = \sqrt{{\sigma_{C_1}}^2 + {\sigma_{C_2}}^2} \quad \text{(Equation 3.5)}$$

The standard deviation of τ_q, expressed as σ_{τ_Q} can be estimated in an analogous manner.

In summary, for the four similarity indices outlined, their between-catchment variabilities across the entire MOPEX database of 428 catchments are characterized by σ_S, σ_A,σ_{τ_P}, and σ_{τ_Q} respectively. To ensure that no one index overwhelms the others by virtue of its numerical scales, the variance of each index, whether it contains all 428 catchments or a smaller subset of them, is normalized by the four constants listed above. For any group of *m* catchments, we define a new quantity, *E*, the metric of regime similarity associated with that group, as follows:

$$E = \sqrt{\left(\frac{\sigma_{S,m}}{\sigma_S}\right)^2 + \left(\frac{\sigma_{A,m}}{\sigma_A}\right)^2 + \left(\frac{\sigma_{\tau_P,m}}{\sigma_{\tau_P}}\right)^2 + \left(\frac{\sigma_{\tau_q,m}}{\sigma_{\tau_q}}\right)^2} \qquad \text{(Equation 3.6)}$$

Essentially, the regime similarity metric, *E*, is a representative measure of the combined within-group variance of the four similarity indices for any group of *m* catchments, with equal weights attached to each of the similarity indices.

3.3.3 Clustering Algorithm: Iterative Dichotomiser Algorithm (ID3)

Classification trees, as discussed in chapter II, offer a straightforward choice for grouping objectives on the basis of similarity or variance measures (Brieman et al, 1984). Such tools are routinely included in many statistical programming packages (Brieman et al, 1993). The clustering algorithm used in this paper is the *iterative dichotomiser* (ID3) algorithm developed by Quinlan (1986), which was re-coded as part of this research.

The algorithm's implementation is explained next. Given all 428 catchments, choose a value of any one of the four indices (e.g., seasonality index) with which to partition all catchments. This will yield two clusters – those with a statistic (i.e., seasonality index) above that value and those with a statistic below that value. Iterate over four possible similarity indices to choose the value of one of these indices that minimizes the regime similarity metric, *E*, and weighted by the number of constituents in each subsequent class (as explained next). Repeat recursively until either value of *E* can no longer decrease significantly or these clusters become too small. The first split, atop the decision tree is offered in detail in the following paragraphs.

By definition, the normalized value of the variance of the entire distribution of any independent variable is equal to unity. Substituting into Eq. 3.6 to obtain the value of *E* of the initial data belonging to 428 catchments yields:

$$E = \sqrt{(1)^2 + (1)^2 + (1)^2 + (1)^2} = \sqrt{4} = 2 \qquad \text{(Equation 3.7)}$$

When all 428 catchments were assessed, although each of the four similarity indices was considered, the best-performing splitting criterion turned out to be a seasonality index of 0.2564. There are 266 catchments with seasonality index values less than 0.2564 and 162 with seasonality indices that exceed 0.2564. The new value of the regime similarity index, *E*, is now calculated as follows:

$$E' = 266(E_{S \leq 0.2564}) + 162(E_{S > 0.2564}) \qquad \text{(Equation 3.8)}$$

In this case, $E_{S\leq 0.2564}$ denotes the regime similarity metric of the set of catchments for which S ≤ 0.2564 (266 in all) and $E_{S>0..2564}$ represents the similarity metric of the set of catchments for which S > 0. 2564 (162 in all). In each term of Eq. 3.8, Eq. 3.6 is now used to estimate *E* for only the sub-groups of 266 and 162 respectively. Substituting into Eq. 3.8 this gives, after the first split:

$$E' = [266(1.1838) + 162(2.1992)]/428 = 1.5681 \quad \text{(Equation 3.9)}$$

This represents the minimum possible value of *E* after one single split. Thus what began with an *E* value of 2 has now improved to 1.5681. It is worth noting that one branch, the one with more seasonal catchments, actually displays greater "disorder" than the entire dataset. However, given that 266 of the 428 catchments begin to cluster significantly (i.e., *E* = 1.1838), the small increase in the disorder of the remaining 162 is justified.

At this point, the algorithm as described above can be repeated recursively, locating an optimal split criterion at each node by choosing from one of the four similarity indices, thus branching outward down the tree. Splitting ceases when it is determined that the catchments within a given terminal node are maximally similar – no further splitting will decrease the regime similarity metric significantly, or there is only a single catchment left at that node (and thus *E* is zero). In some cases, there are very few catchments left in a given node to be split with an obvious pair of clusters. In such cases, adopting different splitting criteria might yield the same two groups. In these rare cases, a manual choice of splitting is invoked to choose the most appropriate class delineator.

3.4 Results: What Patterns Emerge, and Where are the Largest Clusters?

We now present the results of the application of the clustering algorithm presented above, describing the breakdown developing at each level of the decision tree. For presentation purposes, depending on the magnitudes of the similarity indices at which the splits occur, we divide each similarity index into several (3 to 5) distinct and meaningful classes. The combination of these classes then produces the nomenclature we need to describe the catchment classes at each level.

3.4.1 Nomenclature for Catchment Classes

The nomenclature we have adopted is letter-based, using up to 5-letters of the alphabet to characterize the range of values of each of the four similarity indices; these are presented below.

Codes for Aridity Index

V = "Very-Humid," $E_p/P < \sim 0.5$

H = "Humid," $\sim 0.5 < E_p/P < \sim 0.75$

T = "Temperate," $\sim 0.75 < E_p/P < \sim 1.2$

S = "Somewhat Arid," $\sim 1.2 < E_p/P < \sim 2$

A = "Arid," $\sim 2 < E_p/P$

Codes for Seasonality Index

L = "Low Seasonality," $S < \sim 0.25$

I = "Intermediate Seasonality," $\sim 0.25 < S < \sim 0.5$

X = "eXtreme Seasonality," $\sim 0.5 < S$

Codes for Day of Precipitation Peak

J = "June," Max rainfall occurs in early or mid-summer. (Not necessarily in June)

W = "Winter," Max rainfall occurs in winter. (mid-February or March)

B = "Blizzard," Max rainfall in late November. to mid-February)

P = "Printemps," Max rainfall during spring.

Codes for Day of Runoff Peak

Q = Max runoff during summer months. Early June through August.

F = "Fall/Hurricane season." Generally in September/October. Uncommon (TX & FL)

M = "Melt," Spring melt, usually at a peak in April, May, or early June.

C = "Cold runoff," max runoff occurring from early February to before April.

D = "December," max runoff during December/January or early February.

The classes described above can theoretically describe 3 x 5 x 4 x 5 = 300 different combinations of the similarity indices (and associated catchment groupings) although, as will become apparent soon, an overwhelming majority of those combinations will never occur. The nomenclature for these classes was developed after seeing the clusters that emerged. For instance, with respect to the aridity index, there were a few classes where E_p/P was much lower than 0.5, some classes with E_p/P greater than 2.5, and three notable groupings in between. For this reason, five classes were selected. However, with respect to seasonality, in examining groups it became evident that there were catchments with very little seasonality, catchments with extremely high rates of seasonality, and intermediate catchments. Thus, three were chosen. The intention had been to generate as few classes as possible. Indeed we will show that the first 6 most dominant classes will encompass 331 of the 428 catchments.

3.4.2 Initial Split: Top of the Classification Tree

The classification tree begins with the complete database of 428 MOPEX catchments. As mentioned before, the population of 428 catchments is split recursively into smaller, more homogeneous groups, being named along the way depending on the value(s) of the similarity indices at play at each split. After the very first split, the dataset is divided into two large clusters, which are not terminal nodes, but rather, are intermediate nodes, and these are further

split into four clusters, and so on. After each split, the resulting pair of clusters begins to receive a more detailed code using the letters above, depending on the value of the similarity index that is in play at each split.

Seasonality turned out to be the single most important factor in creating order in the 428 catchments in the MOPEX database at the first level. Two clusters emerged: one characterized by catchments with a low seasonality index (L) and the other characterized by catchments with a "not low" seasonality index. This is shown in Fig. 3.6, with the left branch labeled "L" and the right branch labeled "*" because it could be either an "I" or an "X" type of seasonality. The transfer of the first level split onto a map of the United States makes the classes resulting from the first split easy to understand hydrologically, as seen in Fig. 3.7.

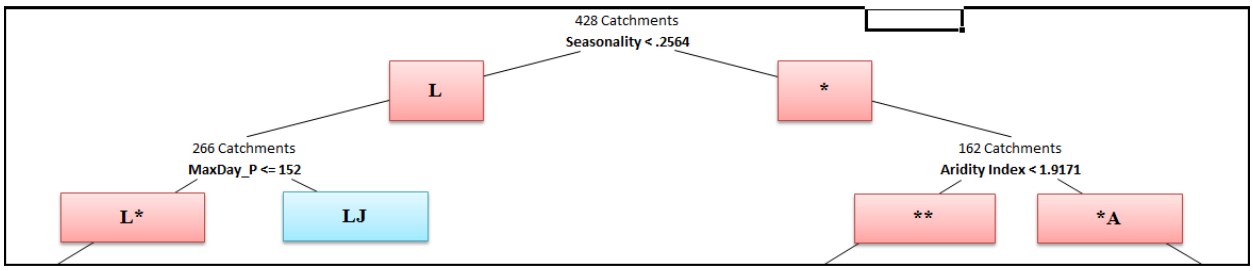


Figure 3.6 – The Top 2 Layers of the Classification Tree (Terminal Node Shown in Blue)

The results presented in Fig. 3.7 indicate that the seasonality index, after only one binary split, effectively partitions the continental United States geographically in a meaningful way. In the eastern part of the country, rainfall is relatively uniform throughout the year, from New England in the northeast, down the Appalachian Mountains to the Ozarks, stretching into the Midwest. Only three eastern catchments in this database, those in Florida, deviate from this pattern, as they see large amounts of rainfall during a warm, humid, hurricane-influenced

summer/autumn and considerably less during the winter. In the western United States, excluding a handful of catchments in the northern Rocky Mountains, every catchment displays considerable seasonal variability of precipitation, from the Midwestern catchments characterized by a precipitation that is in phase with potential evaporation to the Pacific coast catchments in which the precipitation regime is out-of-phase with respect to that of potential evaporation.

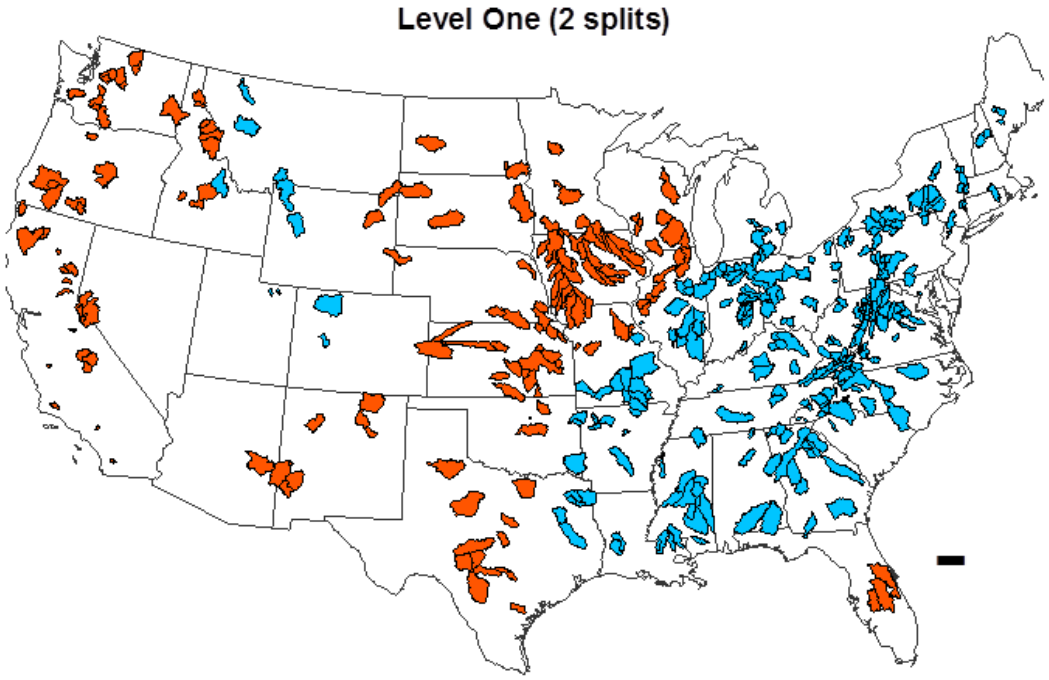


Figure 3.7 – 1st Split, Low Seasonality (Blue) and Higher Seasonality (Orange)

The second split criterion, for the lower seasonality, eastern catchments (colored blue in Fig. 3.7), is the timing (day) of precipitation peak while for the more seasonal, western catchments, the split criterion becomes the aridity index (see Fig. 3.3). For less seasonal catchments, the dividing date falls on June 1st, for the more seasonal catchments, the dividing aridity index is roughly 1.9. This leads to four classes, as shown in Fig. 3.6, one of which (LJ) is

a terminal node. The transfer of these four clusters after two consecutive splits of the original dataset, on to the map of the United States is presented in Fig. 3.8. The results presented in Fig. 3.8 show an east-west division based on the seasonality index at the first level, a northeast-southwest split occurs in the eastern (non-seasonal) region via the timing of rainfall, while in the west, a split based on aridity index distinguishes the Pacific Northwest and the northern Midwest catchments from the remaining western catchments.

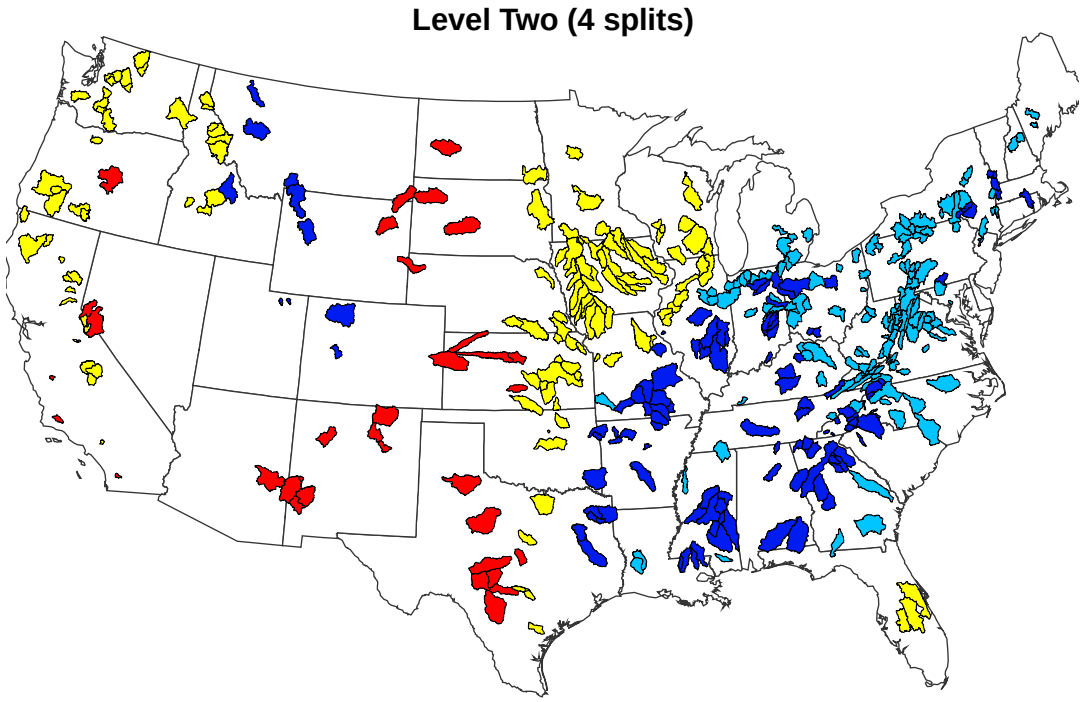


Figure 3.8 – 2nd Split: Low Seasonality & Earlier Precipitation Peak (Dark Blue), Low Seasonality & Later Precipitation Peak (Light Blue), Higher Seasonality and Non-Arid (Yellow), Higher Seasonality and Arid (Red)

3.4.3 The Four Quadrants of the Classification Tree

The four main clusters of similar climatic regions obtained in the second level will be further split by the recursive algorithm outlined above until smaller, very similar climate clusters remain. The details of this are not presented here for reasons of brevity; only the resulting final classification tree is presented. Even here, because of the size of the resulting tree, it is most easily viewed in portions, which we call quadrants, relating to the major clusters formed at the end of the level two splits. In what follows, each quadrant of the classification tree is presented and discussed in detail.

3.4.3.1 First Quadrant: Low Seasonality, Max Precipitation before June 1st

Figure 3.9 presents the expansion of the 1st quadrant. Six climate regions describe the 119 catchments that comprise this group. The most populous group, "LWC" contains 52 catchments, 50 of which are located in the south-eastern states. While this terminal grouping has been obtained without the use of the aridity index, using only seasonality and the timings of precipitation and runoff, the 52 catchments all display $E_p/P < 0.87$, displaying a tight cluster of humid catchments where rainfall and runoff peak in February or March. The second-most populated group is "LPC," containing 29 catchments from the eastern Midwest. Once again, although the aridity index has not been used as a split criterion to obtain this cluster, the 29 catchments have similar E_p/P values, near or slightly below one. This class is distinguished from LWC by virtue of maximum rainfall occurring later in the spring. A third, well-populated cluster is found in 28 "LPM" catchments located in the south-eastern regions of the Midwest where rainfall and runoff both peak during springtime. The "LBMH" catchment in Montana, which seems unusual for its geography given its low seasonality, and humidity ($Ep/P \sim 0.67$), 3 "LBMS" catchments from Colorado and Montana which are similar to the LBMH oddity, only

considerably drier (1.17 < *Ep/P* < 1.66), and 6 "LPQ" catchments, also from the mountain west (Wyoming) where rainfall peaks in the spring instead of the winter, round out this quadrant.

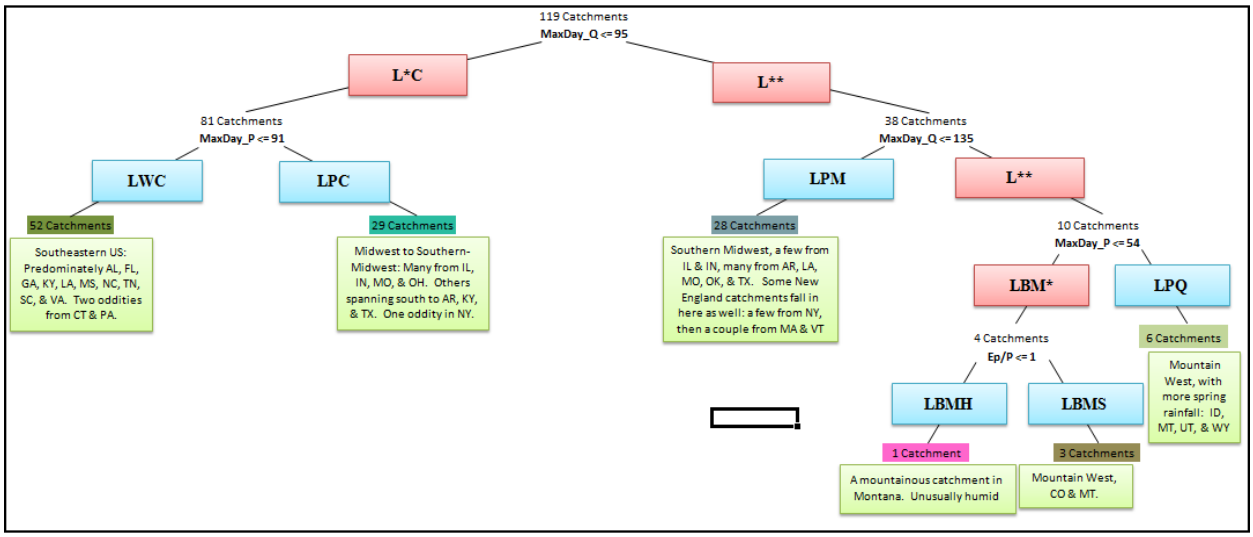


Figure 3.9 – Low Seasonality, Early Precip. Peak: Expanded (Terminal Nodes in Blue)

3.4.3.2 Second Quadrant: Low Seasonality, Max Precipitation after June 1st

This quadrant becomes fully organized with only two criteria for splitting, leaving 147 catchments which all carry the "LJ" designation. Although the maximum date on which runoff occurs is not used to create this class, 145 of the 147 catchments observe maximum runoff between mid-February and late-April (the remaining two peaks occur in the first week of May). In fact, 124 of the 147 catchments peak between the second week of March and the first week of April. Furthermore, although once again the aridity index is not used to generate this cluster, all 65 catchments fall between $E_p/P \sim 0.5$ and $E_p/P \sim 1.05$. This class of catchments defines the mid-Atlantic and Appalachian regions of the United States, extending into the eastern Midwest. This quadrant of the tree, albeit expressed as a single node, is illustrated in Fig. 3.10.

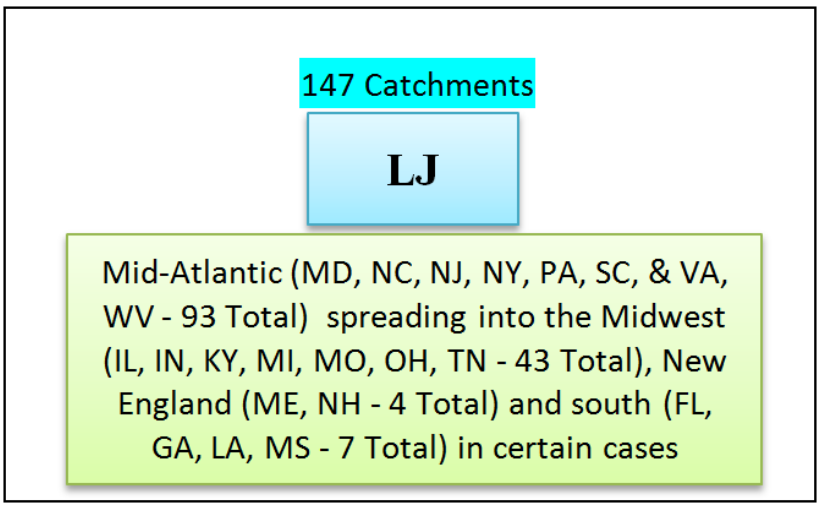


Figure 3.10 – Low Seasonality, Late Precipitation Peak: Expanded (Terminal Node)

3.4.3.3 Third Quadrant: Higher Seasonality, Non-Arid

This quadrant of the tree is the most diverse by a considerable margin. The criteria for this quadrant are, to reiterate, the seasonality index > 0.2564 and an aridity index (*Ep/P*) below 1.9. Some 128 catchments meet these conditions and are further segmented into 12 terminal nodes as shown in Fig. 3.11 (note that two terminal nodes classify to the common Midwestern climate of "ITC"). However, despite the apparent complexity, two climates describe 75 of the 128 catchments. The first, and most common, "ISQJ" contains 39 catchments in the Midwest and southern Midwest. The second most common, "ITC," is quite similar to the previous grouping geographically, although it is more humid and maximum runoff arrives sooner. This class contains 36 catchments from the Midwest and northern Midwest. The remaining clusters, partitioning the Pacific Northwest, consist of 6 humid catchments from northern California and Oregon ("XHD"), 7 catchments in Idaho ("IHM"), 6 extremely humid catchments in Washington ("XVM"), 10 more temperate catchments in the pacific northwest from Washington, ("XTM"), 3 extremely humid catchments in Washington ("IVD"), which differ from their

"XVM" counterparts by virtue of their lower seasonality index, and winter runoff peak, 3 Floridian catchments ("ITF"), which are truly unlike any others in the United States, 7 drier Midwestern catchments with early runoff peaks ("ISCJ"), 3 drier southern Californian catchments ("XSC"), and 6 drier southern Californian/Nevadan catchments with later runoff peaks ("XSMB").

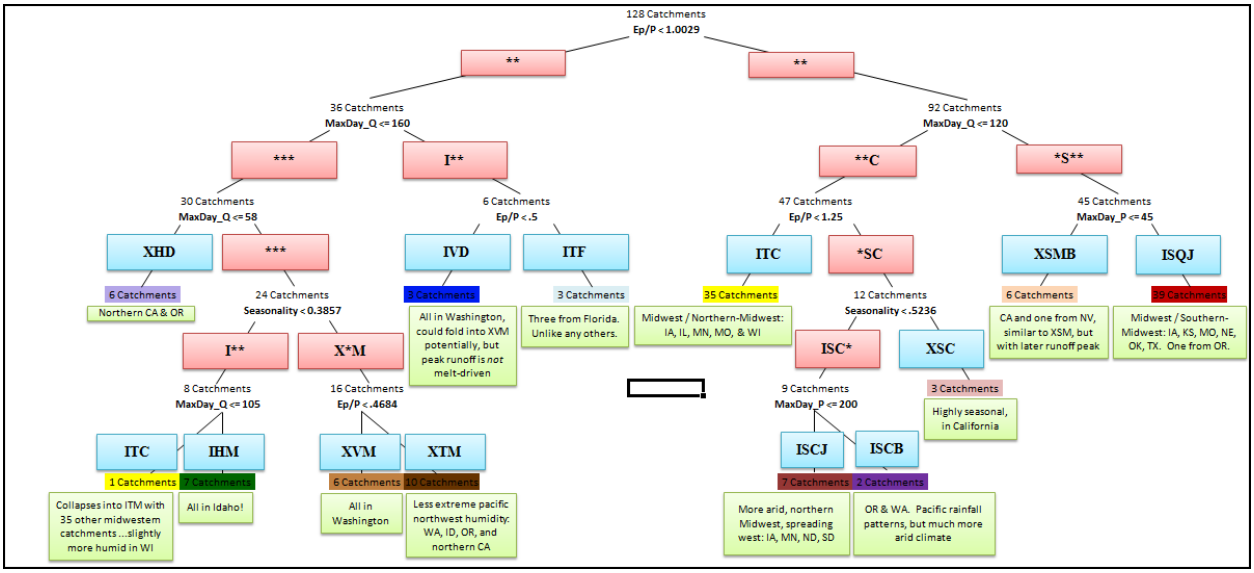


Figure 3.11 – Higher Seasonality, Non-Arid: Expanded (Terminal Nodes in Blue)

The terminal nodes in the 3rd Quadrant contain 10 or fewer catchments, describing certain niche climates of United States. These mini-clusters often describe several catchments that are very similar to each other, but quite different from their neighbors.

3.4.3.4 Fourth Quadrant: Higher Seasonality, Arid

In this final quadrant, the 34 remaining catchments are further divided into five terminal nodes. The most common classification ("IAQ") contains 16 catchments, a miscellaneous assortment of the country's most arid locations, including 10 from the southwest, 5 from the

Midwest, and one remarkably arid catchment in Wyoming (the mountain west). The remaining clusters consist of three Californian catchments that represent the driest American Pacific climates ("XADB"), the northern Midwestern 'badlands,' six extremely arid catchments in Nebraska and North and South Dakota ("XACJ"), a cluster of seven arid southwestern catchments, one oddity in the Pacific northwest, ("IACJ"), shielded from the Pacific coast by the Cascade mountains, and three arid catchments in Texas characterized by runoff peaks occurring as late as the fourth week of October ("IAF"). Fig. 3.12 presents the final quadrant.

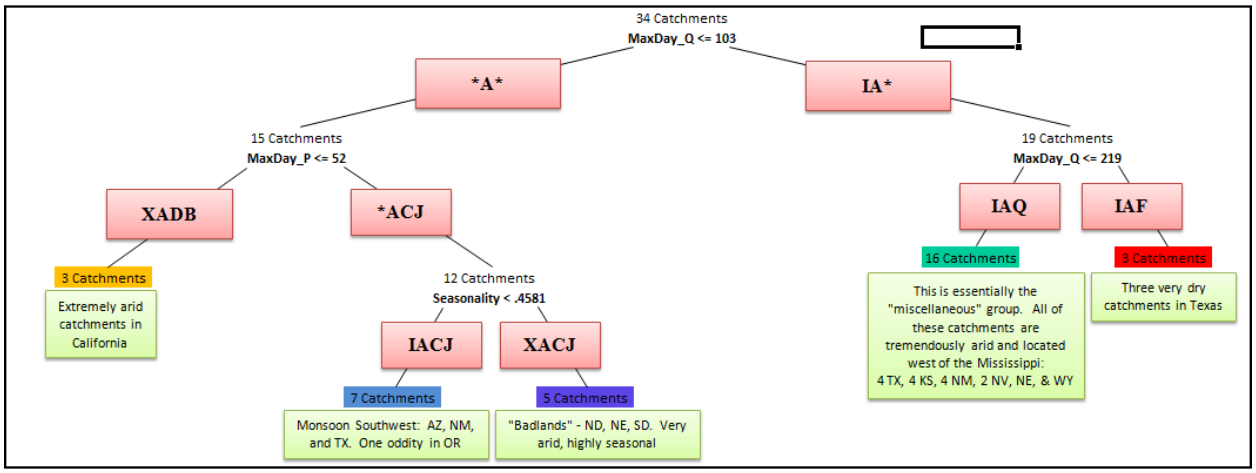


Figure 3.12 – Higher Seasonality, Arid: Expanded (Terminal Nodes in Blue)

3.4.4 Summary of the Resulting Catchment Classification and the Largest Six Classes

In total, the classification tree yielded 24 terminal nodes, depicting 24 distinct classes according to the criteria we have used. The geographic representation of these 24 classes on a map of the continental United States is presented in Fig. 3.13, revealing distinct regional associations of many of the major catchment classes.

However, of the 428 catchments which comprise the MOPEX dataset, 331 can be described by only six climate classes, namely "LWC", "LPC", "LPM", "LJ", "ITC", and "ISQJ".

There are only two other groupings that contain 10 or more catchments. While admittedly, this could be due in part to the makeup of the MOPEX dataset, which contains more catchments in certain regions than in others, it still suggests that while 300 different classifications are theoretically possible using the coding system adopted here, over 77% of the catchments are well described by 2% of the possible classes, and the entire dataset is captured by 8% of all possible classes. In terms of the overall variance of the full dataset, the following are the within-group variances for the six most common classes: LWC – 26.9%, LPC – 23.9%, LPM – 29.8%, LJ – 43.3% (with 140+ catchments), ITC – 28.1%, ISQJ – 46.5%. Considering that these groups comprise 77% of the database, this is quite encouraging, as these clusters contain much less than half of the variance of the original dataset using very simple indices.

Three of the largest six classes are found within the 1st Quadrant (LWC, LPC, and LPM). The catchments belonging to these three classes are characterized by limited seasonality, and are essentially catchments with pre-spring maximum rainfall and runoff (LWC), catchments with pre-spring peak runoff, but mid-Spring maximum rainfall (LPC) and catchments with springtime rainfall and runoff peaks (LPM). The entire set of catchments in the 2nd Quadrant (belonging to the terminal LJ class) is clearly the fourth member of the largest six classes. These catchments display limited seasonality, humid climates, peak runoff during the springtime (likely melt-driven from the Appalachian mountain range), and peak rainfall during the summertime. The final two members of the largest six are found in the 3rd Quadrant, namely ISQJ and ITC (and none of the classes in the 4th Quadrant fall within the largest six). The two Midwestern classes, ISQJ and ITC, both contain catchments with rainfall that is in phase with potential evaporation. However, the ISQJ catchments are notably more arid, with E_p/P averaging roughly 1.5, as opposed to an average of roughly 1 for ITC. As a result of the more temperate climate, the ITC

group displays peak runoff during early spring, when stored water from winter has thawed. On the other hand, ISQJ, characterized by drier soils shows its runoff peak in late-May or June, at the same time as its precipitation peak.

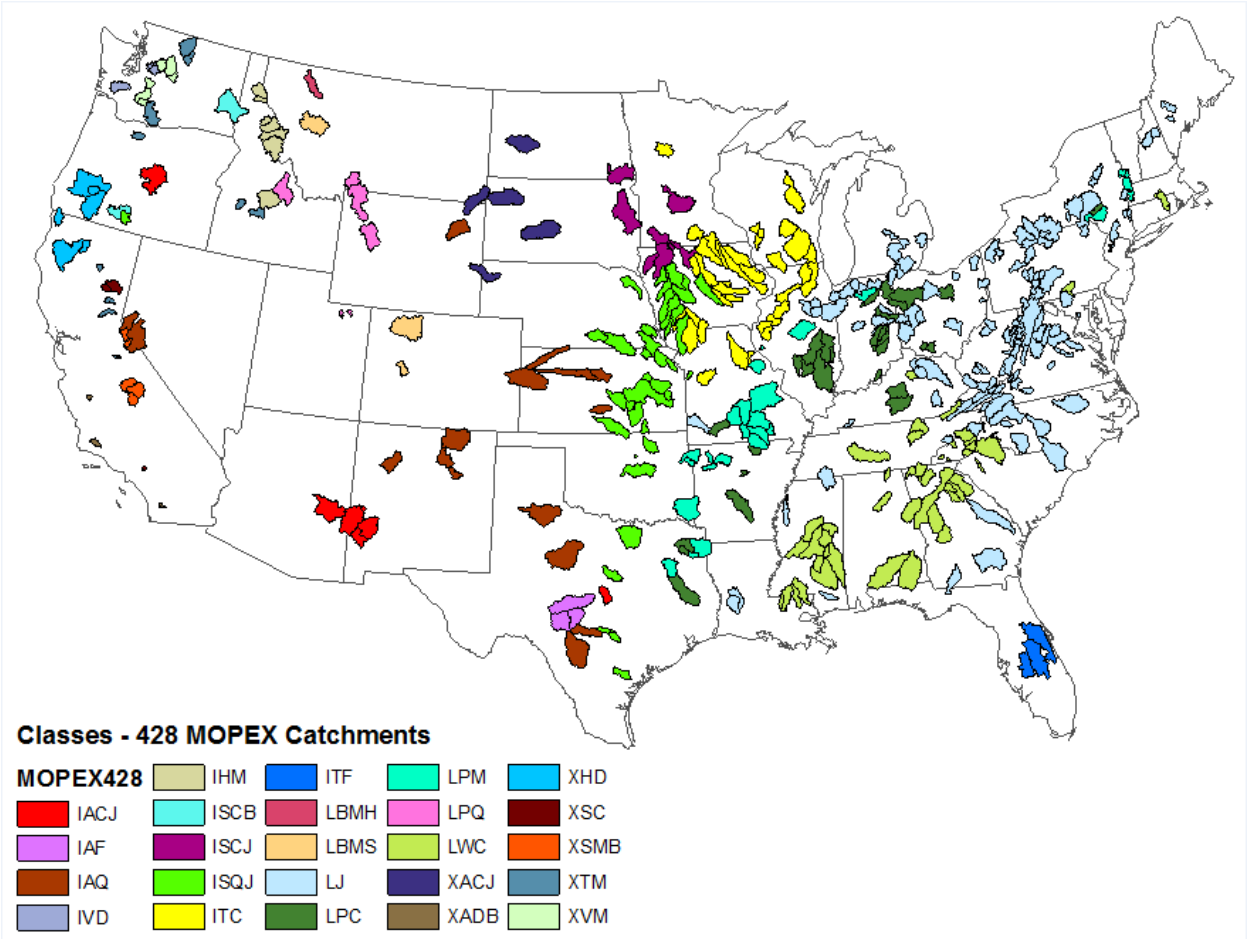


Figure 3.13 –All Classes

3.4.5 Robustness of Classification: Recurrent, Dominant Clusters

When classification systems are generated using recursive, splitting algorithms (those that minimize variance at each stage without concern for future splits), there is a tendency to over-fit one's data. Although variance is minimized at every stage, ensuring that we do not split a group of catchments without purpose, caution is required to prevent fitting the noise inherent in

the dataset rather than true patterns. To this end, the same algorithm was applied to a 197-catchment subset of the larger, 428-catchment dataset. These 197 catchments were chosen due to their comparatively richer datasets (fewer missing days, more complete years, etc) and form the dataset that has been employed by Ye et al. (2012), Cheng et al. (2012), and Yaeger et al. (2012) in the accompanying papers that are all focused on exploring the physical controls of the FDCs from different perspectives.

Although naturally there are subtle differences in the tree that is formed in the latter case, the important features of the tree remain unchanged (not presented for reason of brevity). Using the notion of quadrants, as described in the previous section, the first quadrant is not only characterized by lower seasonality and peak rainfall before June 1st (the same date as the 428-catchment tree), but also contains six total terminal nodes as well. Furthermore, the three most common classes from that quadrant of the 428-catchment tree and their number of constituents ("LWC" – 52, "LPC" – 29, and "LPM" – 28) are mirrored in the most common classes in the 197-catchment tree as well ("LWC" – 21, "LPC" – 11, and "LPMT" - 7). The additional letter T from the subset tree is the result of slightly different sequences of the splitting, yielding essentially, the same groups.

The second quadrant of the 197-catchment tree, like the 428-catchment tree, contains the single class "LJ", now composed of 65 catchments rather than 147. The remaining two quadrants also show some differences, this likely resulting from the nature of the dataset. Unfortunately, given the challenges associated with data gathering in more arid catchments, many of the catchments characterized by substantial missing data are found in the nation's more arid locations. As a result, the 428-catchment tree contains a much higher proportion of arid catchments (thus, the second split criterion, for more seasonal catchments is E_p/P). Conversely,

the 197-catchment tree contains a smaller proportion of arid catchments, and thus, splits on the maximum day of precipitation to define its third and fourth quadrants. Despite this, the two most common classes on this side of the 428-catchment tree ("ISQJ" – 39 and "ITC" – 36), still find their parallels in the 197-catchment tree ("ISQJ" – 15 and "IJTC" – 20). The remaining, less common groups do display some overlap, although in this case differences appear simply because certain groups are not represented at all in the 197-catchment subset, or find themselves folded into other classes. Despite these minor distinctions, once again, a nearly identical largest six again defines over 70% of all catchments and the tree's general structure remains intact. This demonstrates that not only is the classification system intuitively satisfying in its simplicity, but is robust to alterations in the dataset.

The most effective argument for the success of this classification system lies in its ability to validate the initial hypothesis – that simple climatic regime indicators lead to clusters of similar runoff behavior. Each cluster of runoff regimes is presented in Fig. 3.14, demonstrating regional self-similarity. While certain clusters with larger numbers of catchments (notably LJ) do display some variance among their constituents, the overall pattern of runoff timing associated with each catchment still remains intact. Moreover, analysis of the flow duration curves by class confirms our initial speculation, as these flow duration curves are organized by a classification tree constructed solely with regime curve features. Table 1 quantifies the decrease in variance with respect to 100 key percentiles of the FDC as one progresses down the classification tree. In other words, not only are the four key indices being grouped effectively, but the FDC's of the constituent groups are well-organized as well. More detailed discussions of this connection and its relationship to other findings from the first two papers of this series can be found in Yaeger et al (2012).

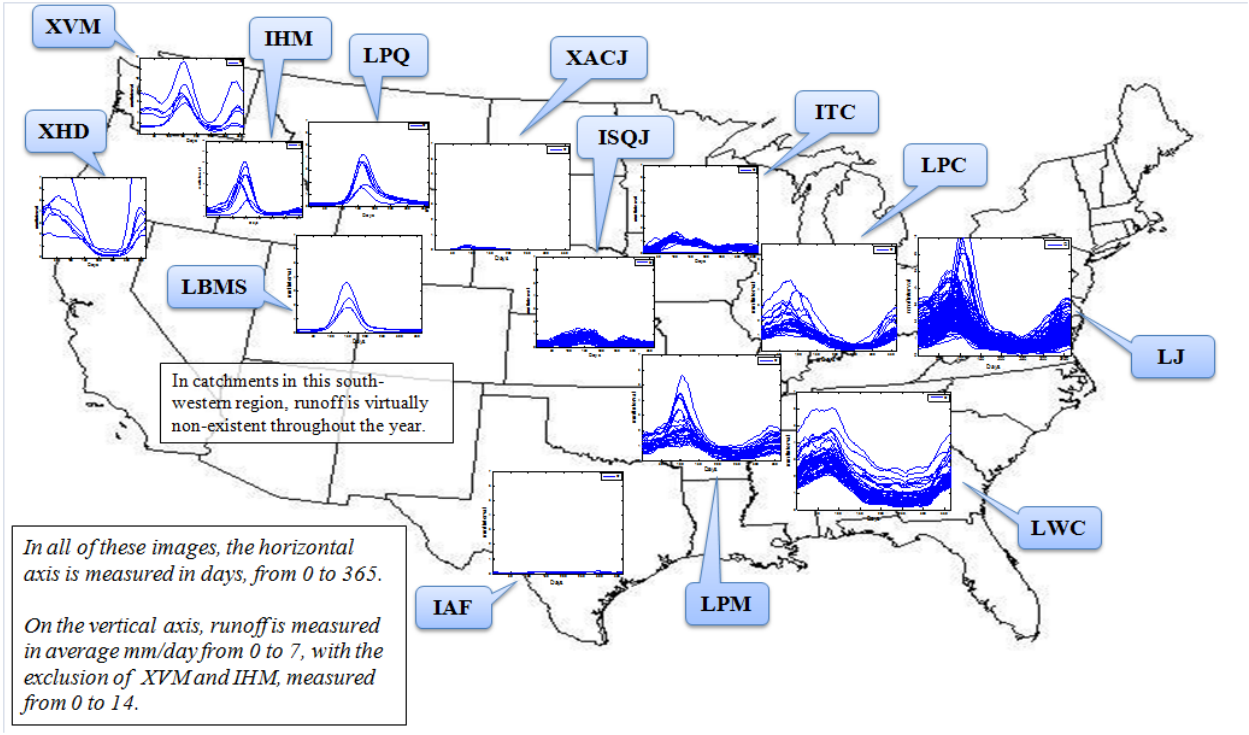


Fig. 3.14: Clusters of Runoff Regimes

Level 0	100%
Level 1	86%
Level 2	68%
Level 3	63%
Level 4	54%
Level 5	47%
Level 6	47%
Level 7	46%

Table 3.1: Decreasing FDC variance (layer-by-layer down the tree)

3.5 Conclusion: What is Learned and Unanswered Questions

This chapter has presented the application of a clustering algorithm (i.e., Iterative Dichotomizer or ID3 algorithm) for classifying catchments across the continental United States

with respect to their climatic seasonality and regime behavior (i.e., mean within-year variation of runoff). The classification was achieved by assessing the catchments in terms of a metric of regime similarity, E, which is a composite measure estimated on the basis of the magnitudes of four similarity indices: (i) a seasonality index of precipitation, (ii) aridity index, (iii) timing of maximum precipitation, and (iv) timing of maximum runoff. The clustering algorithm was applied to 428 catchments across the continental United States belonging to the MOPEX dataset.

The clustering algorithm identified 24 distinct classes. Even though the classification was achieved with just 4 numbers from each catchment (similarity indices), and only the max date of the runoff regime curve was used, the regime behavior for each of the classes showed distinct differences between classes and strong similarity within. This confirms the power of the simple classification scheme for predicting regime behavior across the continental United States, subject to the limitations of the geographical extent of the dataset and coverage across the country. Considering that three of the four indices used to construct the classification tree are based upon climate, it comes as no surprise that climate's impact is readily apparent. Just as Koppen-Geiger delineated the nation into clusters of climatic similarity a century ago, climate still dominates the hydrologic landscape, creating distinct, hydrologically similar clusters.

The resulting classes also display strong regional associations and patterns, which is very valuable to further explore the climatic and landscape controls underlying the resulting catchment classes. Whether the final group is IHM, with seven catchments all located in the state of Idaho or ITF, with three catchments all located in Florida, or IAF, with three catchments located in one region of Texas, these groups are not only numerically similar, but geographically contiguous as well in many cases.

Despite the enormous heterogeneity of catchments represented in the MOPEX dataset, just six classes accounted for over 77% of the catchments. The classification system is found to be robust, producing the same recurring six clusters even with a smaller subset of the full dataset. Each of the recurring, dominant six classes display distinct characteristics that suggest their own set of hydrologic drivers. In the Midwest, the aridity index determines whether runoff is driven by the spring thaw of water frozen in soil storage (ITC) or by the arrival of summer rains (ISQJ). In the south-east, runoff timing in late-winter (LWC and LPC) or early-Spring is governed by the temperatures in and around the Applachian Mountains. In the north-east, spring runoff is likely the result of melting of snow (LJ). In the area in which the south-eastern United States merges with the Midwest, seasonality begins to appear strongly, and runoff is driven by springtime rainfall (LPM). In addition to these largest six clusters, other smaller, niche climates display distinct behaviors. From the monsoon-driven southwest (XACJ), where precipitation occurs mainly in a narrow band of summer months, to the extremely humid Pacific Northwest where runoff peaks are driven by extreme winter rainfall (IVD) or the melting of snowcaps in the spring (XVM), the United States exhibits a tremendously heterogeneous group of catchments.

The analyses presented in this paper have identified catchment groupings that are similar in terms of their runoff regime. What makes them similar? Their regime curves certainly suggest as much, but does that imply similar dominant processes? The accompanying paper by Ye et al. (2012) explores their regime behavior from a process perspective, by adopting a top-down modeling approach. Is there a recognizable mapping between the catchment classification found in this paper and the classification of dominant processes highlighted in Ye et al. (2012)? Furthermore, this paper has been motivated by our quest to explore the physical controls of the Flow Duration Curve (FDC), considering that the regime curve provides a major connective

tissue between the high flow and low flow ends of the FDC. Cheng et al. (2012) presented an empirical analysis of the regional patterns of FDCs across the continental United States, and their physical controls. Is there a connection between the regional groupings of catchments based on the regime curve, and regional patterns of variation of the FDCs? The accompanying synthesis paper by Yaeger et al. (2012) addresses these questions through cross comparisons between the results of each of these three studies to draw general conclusions about the physical and process controls of the regime curve and the Flow Duration Curve, helping to discover not only which catchments are similar, but also *why* they are similar.

Having established these climate classifications, the next step is to study how these hydroclimatic regimes may shift over time, causing the classifications to change. The next chapter addresses this issue, offering a methodology for analysis of hydroclimatic change.

CHAPTER 4

PATTERNS OF HYDROCLIMATIC SHIFTS: AN ANALYSIS OF CHANGING HYDROCLIMATIC REGIMES

Temporal shifts in rainfall and runoff regime curves appear throughout the continental United States, but differ from region to region. This chapter explores these regime shifts by building upon the hydroclimatic classification system presented in the previous chapter. The same MOPEX data from over four hundred catchments during a 55-year period are analyzed to reveal how the indicators have shifted before and after 1970, before and after 1975, and before and after 1980. Statistically significant hydroclimatic changes in these indicators are explored qualitatively, suggesting which catchments today might resemble other catchments tomorrow. Thus, a preview of current locations in one class under future conditions is provided by observing existing locations of another class. The classification system structure enables organization of these data, allowing patterns of regime change to emerge without highly-specified models at each individual site. Regional analyses define changes in mean seasonal rainfall/runoff regimes, including shifts in the daily variability of rainfall and runoff.

Additionally, changes in regime curves of minimum and maximum rainfall/runoff observations are analyzed and discussed. Results indicate that after 1980, classifications typically found in the southeastern quarter of the U.S. have expanded northward and westward. Regionally, the Midwest and Rocky Mountains seem to demonstrate more frequent, but less intense storms after 1980, while southeastern catchments are shown to be processing much less water in the form of precipitation and runoff than in previous years.

4.1 Introduction and Objectives

Chapter III classified 428 catchments throughout the continental United States on the basis of four hydroclimatic indices derived from 55 years of daily precipitation, streamflow, and evapotranspiration data. The four indices represent the dominant hydrologic control processes in many United States catchments (Ye et al., 2012). It is this classification system that forms the basis of the analysis presented in this chapter, which explores the shifts of hydroclimatic regime curves, both by region and by the climate classes generated in the national-scale classification system discussed in Chapter III. Comparing the regime curves from the earlier decades of the database to the regime curves formed from the remaining years reveals insight into hydroclimatic trends and the types of functional changes that might inform decision support modeling. In some cases, the classification of a given catchment may have changed, requiring either a recalibration of the machine learning model or, alternatively, the repurposing of an existing model for a different class. By understanding large-scale spatial and temporal variability, more informed decision-support should be enabled at any location at any point in time.

This chapter begins with a description of our methodology for assessing hydroclimatic shifts over the 55-years of data available within the MOPEX database, how hydrologic class "shifts" are determined, and how significance in these changes is ultimately established. The analysis focuses on alterations in the distributions of precipitation and runoff throughout the year during the period of analysis – demonstrating where non-stationarity is most notable in the United States.

The work then discusses the "regional stories," comparing and contrasting the impacts of hydroclimatic shifts in various regions, both in terms of the classes determined by the classification system and by groups of classes that fully specify a region of the United States,

such as "the Midwest." It will become apparent that while hydroclimatic shifts are a national phenomenon, the specific impacts vary greatly by region both in terms of magnitude, variability, and in terms of annual timing. Implications for future modeling and decision analysis are then discussed, particularly when a given location's hydroclimatology in fifty years may well resemble today's hydroclimatology of a completely different location. Ultimately, the objective of this chapter lies in presenting the empirical hydroclimatic changes over fifty-five years of rainfall and runoff data on a continental and regional level. Use of a classification system to organize over four hundred catchments allows, even without highly-specific models at each location, the identification of hydroclimatic trends throughout the United States.

4.2 Methodology

The climate classification systems described in Chapter III partitions the U.S. into 24 clusters, based upon four indicators of hydroclimatic similarity: seasonality of precipitation, aridity index, maximum precipitation day, and maximum runoff date. These four indicators are calculated using daily data from 1948 to 2003. These features were selected as the minimum amount of information required to explain substantial hydrologic diversity. Other features, such as runoff/precipitation ratio, were also examined but ultimately eliminated due to their inability to improve the model and/or their correlation with features already employed.

To analyze hydroclimatic shifts in these classes, the MOPEX data were partitioned into "before and after" analyses during a 55-year period (1948-2003). This period exhibited significant change in terms of global climate shifts, long-term climate cycles, land use change, and human usage of water resources. The "before" and "after" datasets were selected with 1970, 1975, and 1980 as possible years on which to divide the data, ensuring that 20+ years of data

exist in both the "before" and "after" dataset. As a result, any climatic cycle with a period measured in years rather than decades should bear limited influence on the subsequent analysis. However, several important climatic cycles, most notably the El Nino/Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO) and the Atlantic Multidecadal Oscillation (AMO) occur over periods ranging from a few years to decades. Their impacts on this analysis are addressed in subsequent sections.

Note that catchments chosen for the MOPEX project have been demonstrated to show evidence of a changing hydroclimatology (Wang and Hejazi, 2006) and are generally assumed to be unaffected by anthropogenic impacts (Wang and Hejazi, 2011; Duan et al, 2006), which enables discovery of hydroclimatic insights with minimal confounding from anthropogenic influence.

4.2.1 Division and Re-Classification: Before & After 1970, 1975, and 1980

If the 55-year MOPEX dataset includes shifts in hydroclimatic regimes, the classification of a given catchment may change as the decades elapse. Thus, it is useful to divide these 55 years into "before and after" time segments, allowing comparisons of the hydroclimates earlier in the 55-year window with those that more nearly approach the present. For the purpose of robustness as well as additional insight, these analyses were performed using three divisions at 1970, 1975, and 1980. More specifically, a dataset was first created using only time series data before 1970 with its complement containing only data after 1970. The two datasets were then compared in terms of catchment classifications using the original (1948-2003) classification tree (Figure 3.13) and the values of the indices used for classification (e.g., the change in the date of

maximum runoff). This analysis was then repeated for datasets divided at 1975 and 1980, respectively.

This approach allows observation of which division displays the most prominent shifts, providing an assessment of key periods of hydroclimatic changes, the sequence of events, and whether these impacts occurred, on balance, earlier or later in the 55-year window. Unfortunately, many catchments within the MOPEX database are missing data for windows of five to ten years. These divisions (1970, 1975, 1980) were chosen to ensure a sufficiently rich dataset on both sides of the division, and not for any social or geopolitical rationale. The full dataset begins in 1948 and ends in 2003. These dividing years ensure that we have a minimum of 22 years on both sides of the threshold, thus a handful of missing years will not detract from modeling annual patterns.

4.2.2 Shifts in the Four Distributions

Using these three pairs of datasets, the four hydrologic indices (seasonality, aridity index, maximum date of precipitation, maximum date of runoff) were recalculated for each catchment using only data before and after the division year (say 1970). The change in the average value of each index before and after the threshold year (both direction and magnitude) by class yields regional insights into the effects of hydroclimatic change. The three different division years provide insights into the timing of these shifts. Changes in the four key indices may also imply new classifications in the original tree, allowing an observation of how the overall distribution of classes has changed. Moreover, if hydroclimatic shifts cause specific, predictable changes to the four indices, one can then see which locations today are currently most similar to the forecasted

hydroclimate of tomorrow. In this vein, a preview of hydroclimatic conditions in one location might be readily available in another location.

To address impacts of shifts in long-term annual climate indices (ENSO, AMO, PDO), these indices were partitioned into samples before and after the same years (1948-1980 & 1981-2003, e.g.) and then analyzed using the same two-sample, heteroscedastic t-test discussed below. The PDO does display significantly larger values after 1970, 1975, or 1980. However, the ENSO and AMO do not display statistically significant differences at $\alpha = 0.1$ using 1980 as a dividing year. Therefore only the impacts of shifts in the PDO are considered in the results and discussion below.

4.2.3 Tests of Significance

For each threshold year (1970, 1975, or 1980) and hydrologic index, two samples are produced that can be analyzed for significant differences in their mean values using a two-tailed, two-sampled, unequal variance (heteroscedastic) t-test. This test is chosen under conservative assumptions that variables can increase or decrease in value after the given threshold year and that both samples are non-normally distributed with unequal variances. The generalized equation for a two-sampled, two-tailed, heteroscedastic t-test is:

$$t = \frac{\overline{X_1} - \overline{X_2}}{s_{\overline{X_1} - \overline{X_2}}} \qquad \text{(Equation 4.1)}$$

where $\overline{X_1}$ and X_2 represent the means of the two samples, and the unbiased estimate of the variance is calculated as:

$$s_{\overline{X_1}-\overline{X_2}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{(Equation 4.2)}$$

where s_1^2 and s_2^2 denote the unbiased estimates of the variance of the two samples and n_1 and n_2 represent the sample sizes (Welch, 1947), which in this case are equal, as the calculations address the same *n* catchments before and after a given year. While proximal catchments may experience similar effects (and thus are less than wholly independent), the statistical significances observed generally present near-zero p-values. This suggests that statistical dependencies among proximal catchments are less important, as the results would be significant even with smaller sample sizes.

For seasonality, a scalar value, application of this test is straightforward; for variables that represent a date on a circular scale from 0 to 365, the test must be modified. For example, day 3 (January 3rd) and day 363 (December 29th) – only five days earlier, are not significantly different in time. For such cases of circular data, the terms s_1^2 and s_2^2 are calculated using a circular mapping of each date value d scaled from 0 to 365 as follows:

$$C_1 = sin\left(\frac{d}{365}2\pi\right), C_2 = cos\left(\frac{d}{365}2\pi\right) \quad \text{(Equation 4.3)}$$

By estimating s_{C_1} and s_{C_2}, the sample standard deviations of C_1, and C_2, respectively, the standard deviation of d, s_d, can be estimated as:

$$s_d = \sqrt{{s_{C_1}}^2 + {s_{C_2}}^2} \quad \text{(Equation 4.4)}$$

The sample standard deviation s_d can then be used as in a traditional two-sample t-test (Equations 1 and 2). While this manipulation removes directional information (i.e., whether the shifts are positive or negative), the samples' "similarity" is unchanged whether a given date is n-days prior or n-days posterior to a given datum.

4.3 Results – Shifts in Distributions

4.3.1 Seasonality: Increasing and Decreasing Variability of Precipitation Regime Curves

Figures 4.1a, 4.1b, and 4.1c present the changes in seasonality index of each hydrologic class for 1970, 1975, and 1980 cutoff years, respectively. These images show catchments that previously had very limited seasonal variance (receiving approximately constant amounts of rainfall throughout the year) and now receive rainfall much more inconsistently throughout the year (red dots). For example, in the northeastern states, from the Great Lakes through New England, precipitation has become increasingly seasonal, with this effect becoming more pronounced as the cutoff year moves from 1970 to 1980. This same phenomenon is observed within the lower reaches of the Mississippi River. Catchments experiencing less seasonal variance in rainfall are presented as darker blue dots. Though the AMO analysis predicts more significant differences using 1970 as a division than 1980, the opposite impact is observed, suggesting that hydroclimatic cycles do not fully explain the observations.

The northeast experiences an increase in seasonality driven by rainfall arriving earlier in the year than had previously been the case. Conversely, in the southern/Midwestern regions, increases in seasonality are driven by rainfall's arrival later in the season, perhaps due to increased incidence of hurricanes and severe storms from warmer temperatures (Goldenberg et

al, 2001; Anthes et al, 2006). Finally, the southeastern United States observes a flattening of the annual precipitation regime curve, as evidenced by the decrease in seasonality index (Figures 4.1a – 4.1c).

Also noteworthy is the shift towards increasing seasonality in class LJ (Figure 3.13), which is the largest of all of the clusters. This class is characterized by lower seasonality of precipitation (L) and maximum rainfall after June 1st (J). Class LJ comprises a full quadrant of the classification tree and encompasses over 30% of the catchments within the database. This becomes evident when dividing on the year 1980 and is statistically significant ($p \sim 0.02$). However, when the dividing year is 1970, this class shows the opposite impact, displaying a flattening of the regime curve after the division with statistical significance ($p \sim 0.004$). One explanation for this observation could be that the first impacts of a different hydroclimatic regime in this region are simply an increase in fall precipitation, the time of year in which the northeast receives slightly less rain than in other seasons – this flattens the regime curve. However, in subsequent decades, this increase occurs in spring, increasing seasonal variance once again. Visually, the precipitation regime curve flattens initially, but in more recent years, inconsistent, more seasonally-driven rainfall has become the norm.

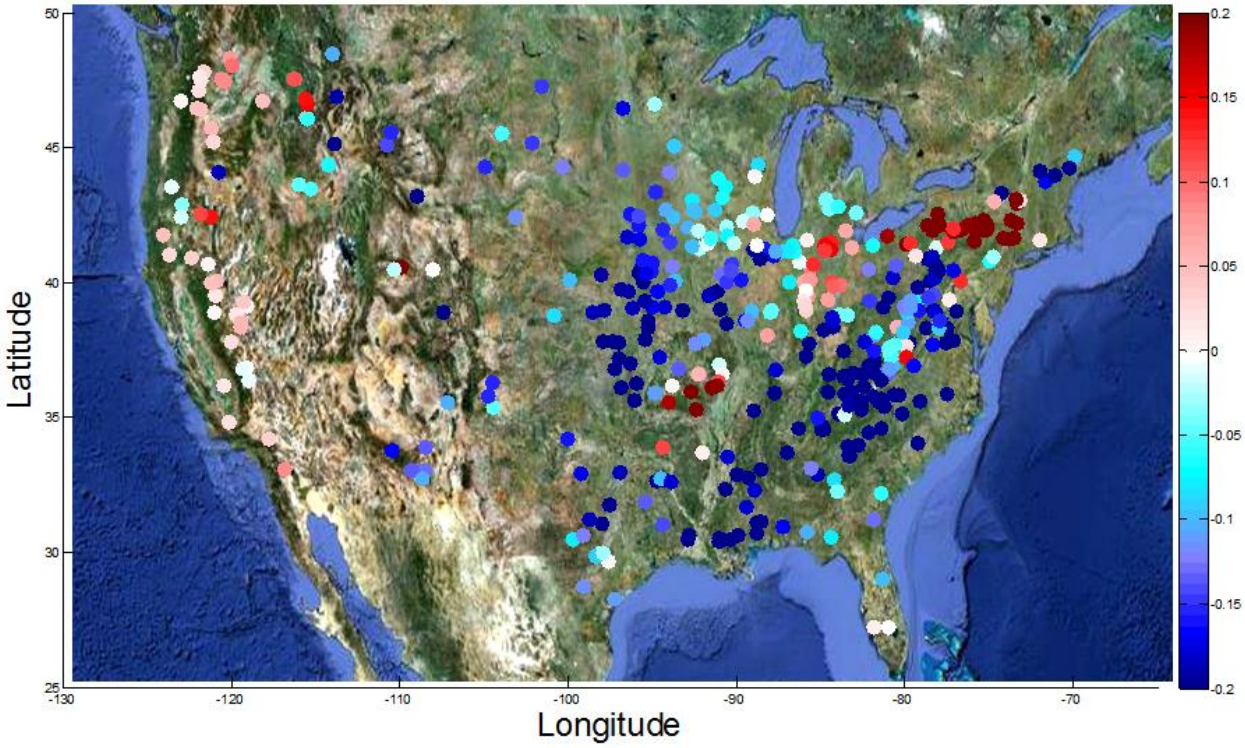


Figure 4.1a – Pre/Post 1970 change in seasonality (0.1 implies a 10% increase)

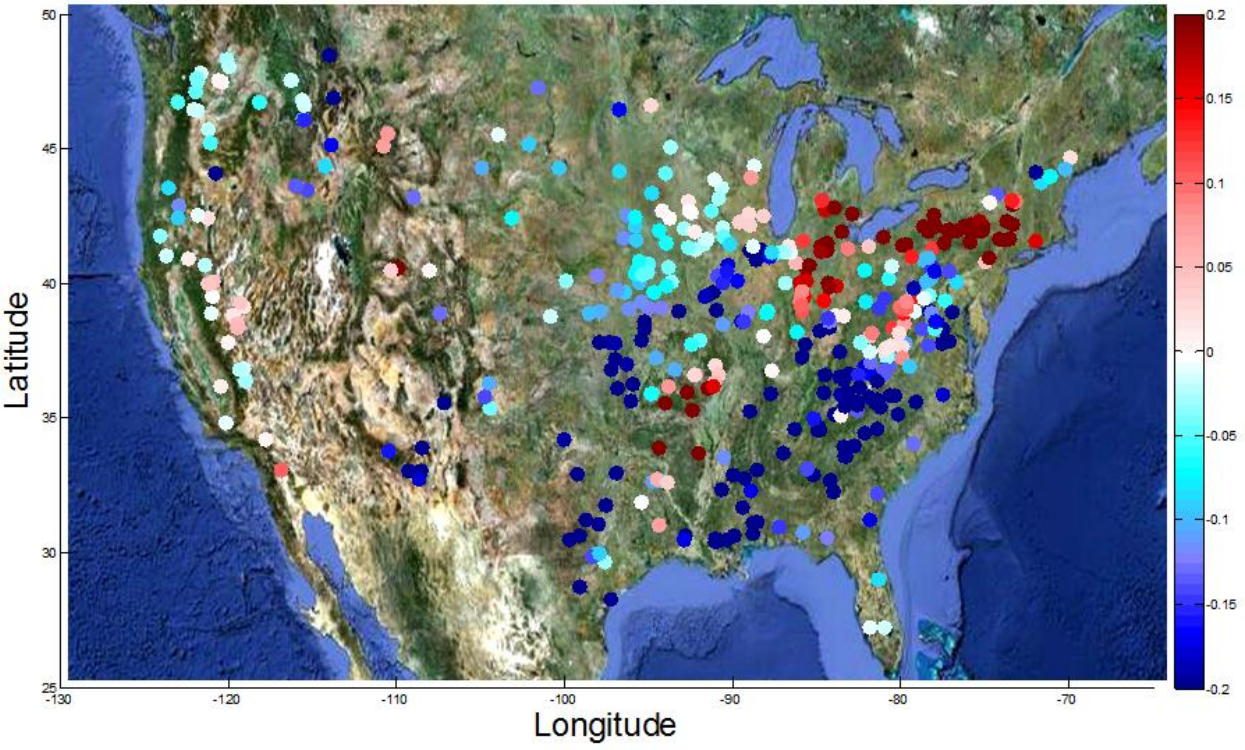


Figure 4.1b – Pre/Post 1975 change in seasonality (0.1 implies a 10% increase)

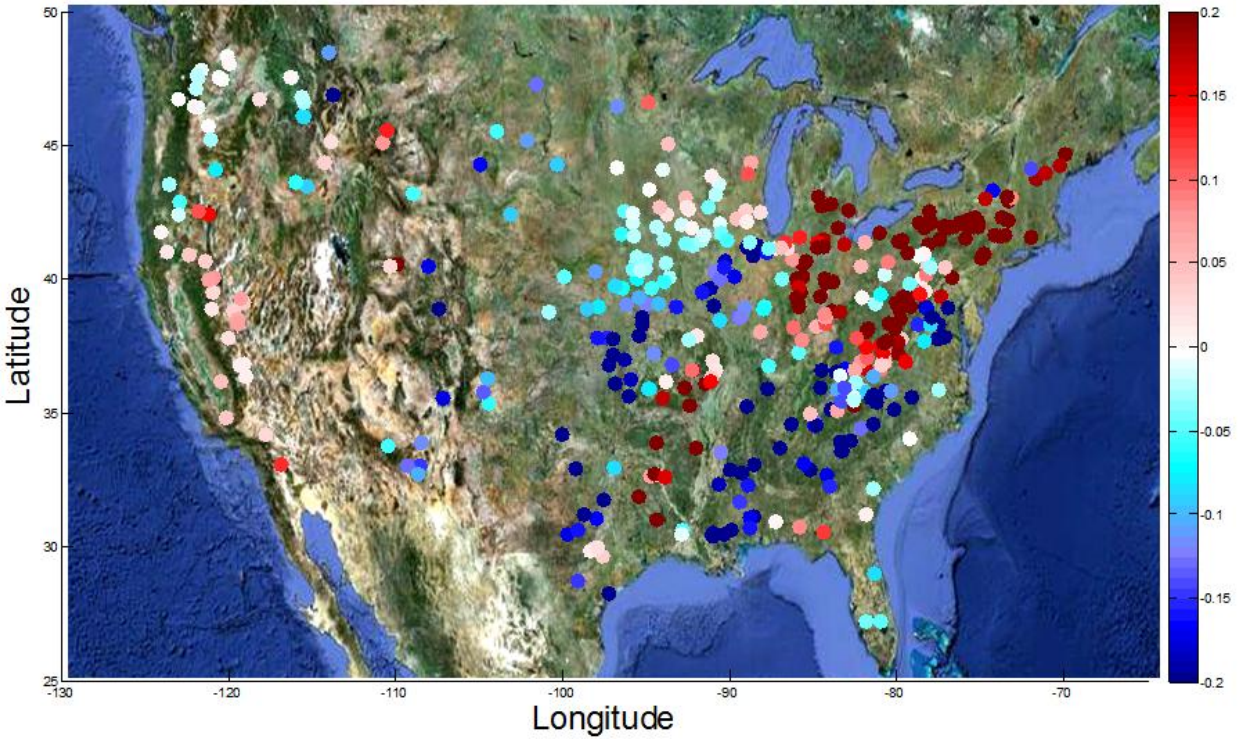


Figure 4.1c – Pre/Post 1980 change in seasonality (0.1 implies a 10% increase)

4.3.2 Day of Maximum Precipitation: Shifts in Precipitation Timing

Table 4.1 presents the shift in the date of maximum precipitation (top number in each cell) and its statistical significance (p-value and bottom number in each cell) for each of the four quadrants (Figure 3.8) of the original classification tree, partitioned on the years 1970, 1975, and 1980. These four quadrants represent the first two bifurcations of the original dataset on the classification tree. The first quadrant contains catchments with lesser precipitation seasonality and maximum rainfall occurring before June 1st. The 2nd quadrant contains low-seasonality catchments with maximum rainfall occurring after June 1st. The 3rd quadrant consists of catchments characterized by higher seasonality, but non-arid climates. The 4th and final quadrant contains catchments that are highly seasonal and highly arid. Note that "-12" for the shift implies that after the threshold year, the maximum day of precipitation arrived twelve days

sooner than it had before the threshold. Results with a significance level of $\alpha = 0.05$ are bolded and italicized.

Shift in date of maximum precipitation	Before / After 1970	Before / After 1975	Before / After 1980
Q1 – Low seasonality, earlier precipitation peak (southeastern U.S.)	**-12** ***(p < 0.001)***	**-11** ***(p < 0.001)***	**-19** ***(p < 0.001)***
Q2 – Low seasonality, later precipitation peak (mid-Atlantic & northeastern U.S.)	**-10** *(p = 0.631)*	**-7** ***(p < 0.001)***	**-24** ***(p < 0.001)***
Q3 – Higher seasonality, non-arid (northern Midwest & Pacific coast)	**-6** ***(p = 0.050)***	**+15** *(p = 0.321)*	**+10** ***(p = 0.004)***
Q4 – Higher seasonality, arid (southwest and drier Midwest)	**-21** ***(p = < 0.001)***	**-17** ***(p = < 0.001)***	**-2** ***(p = 0.015)***

Table 4.1 – Shifts in date of maximum precipitation over threshold years for the four quadrants of the classification tree by Coopersmith et al (2012), with p-values from a two-sample, two-tailed, heteroscedastic t-test. (Bold indicates significant changes; Negative values in red indicate decreases; Quadrants as in Fig. 3.8)

The maximum day of precipitation based on 1970 division shows some initial evidence of shifted rainfall timing, but by the 1980 cutoff (Figure 4), not only are the shifts significant in all four quadrants of the original classification tree but the *least* significant p-value is a scant 0.015 for the 4th quadrant, which contains only 34 of the 428 catchments. The remaining three p-values fall well below 0.01. This suggests that shifts in rainfall timing are meaningful and have become even more substantial in recent years.

Figure 4.2 shows rainfall peak shifts (measured in days) that reveal regional clusters. In the Pacific Southwest, observe that rainfall peaks have shifted forward by approximately one month. A similar phenomenon of forward-shifted rainfall peaks is noted throughout the Midwest, becoming even more extreme near the Gulf of Mexico. As noted previously in terms of increased seasonality, this may be driven by more prominent late-year tropical storms.

Conversely, on the east coast, rainfall is now arriving earlier rather than later, especially in New England where increases in seasonality were observed. One hypothesized mechanism for this phenomenon could be the result of higher spring temperatures leading to more rapid evaporation of snow, which in turn leads to earlier, more voluminous rainfall (Trenberth, 2011). In the Rocky Mountains, where rainfall is rather consistent year-round, the peak date of rainfall has shifted from winter to late spring during snow cap melting. This is likely due to the same phenomenon noted in New England.

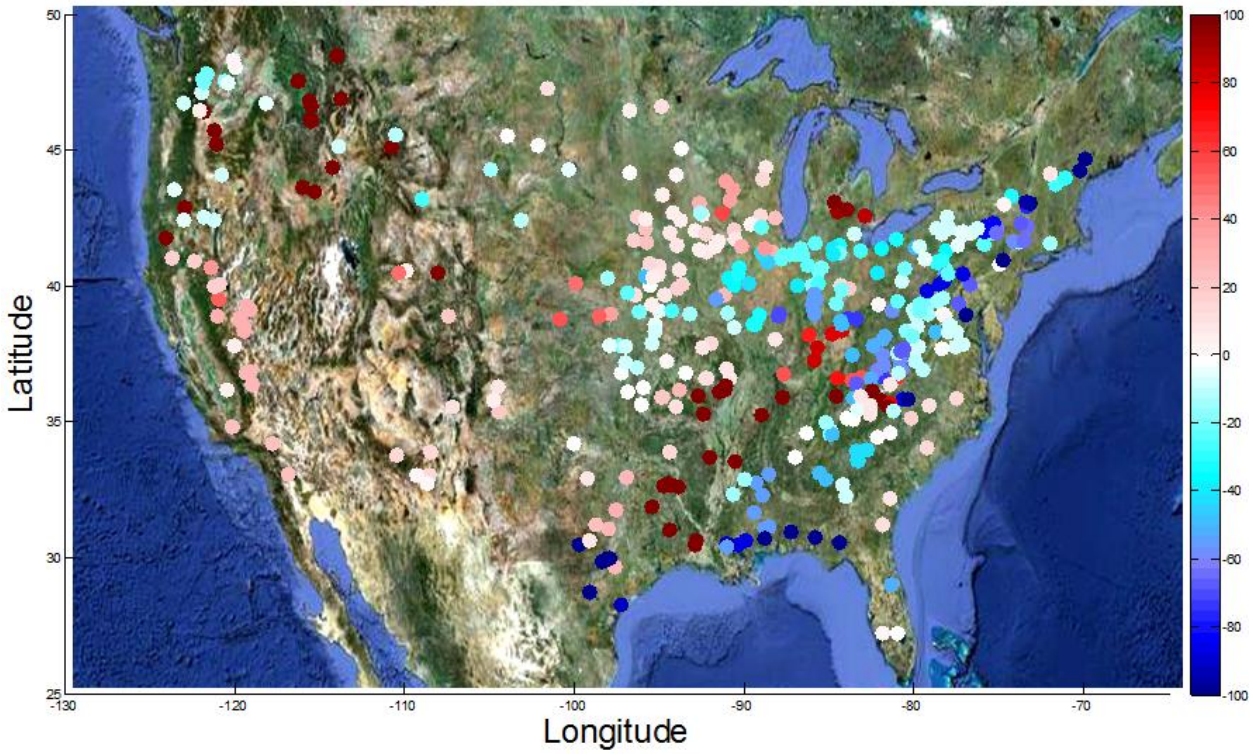


Figure 4.2 – Shift in Annual Precipitation Peak Date (in days), Pre/Post 1980. Color scale ranges from +100 (shifted forward, March to June, e.g.) to -100 (shifted backwards in time, June to March, e.g.)

4.3.3 Runoff Timing: Changes in the Drivers of Streamflow

Table 4.2, like Table 4.1, presents the four quadrants of the original classification, partitioned on the years 1970, 1975, and 1980. The top number in each cell denotes the shift in

the date of maximum runoff. The lower number in the cell is the p-value explaining the statistical significance of that result. As in Table 4.1, results that are significant at $\alpha = 0.05$ are bolded and italicized.

Shift in date of maximum runoff	Before / After 1970	Before / After 1975	Before / After 1980
Q1 – Low seasonality, earlier precipitation peak (southeastern U.S.)	**-1** ***(p < 0.001)***	**-7** ***(p < 0.001)***	**-8** ***(p < 0.001)***
Q2 – Low seasonality, later precipitation peak (mid-Atlantic & northeastern U.S.)	**-3** ***(p = 0.044)***	**-3** ***(p < 0.001)***	**+3** ***(p < 0.001)***
Q3 – Higher seasonality, non-arid (northern Midwest & Pacific coast)	**+8** ***(p = 0.049)***	**+12** (*p* = 0.422)	**+16** ***(p < 0.001)***
Q4 – Higher seasonality, arid (southwest and drier Midwest)	**-9** ***(p = 0.006)***	**-11** (*p* = 0.160)	**-5** ***(p = 0.003)***

Table 4.2 – Shifts in date of maximum runoff over threshold years for the four quadrants of the classification tree in Coopersmith et al (2012) with p-values from a two-sample, two-tailed, heteroscedastic t-test. (Bold indicates significant changes; Negative values in red indicate decreases; Quadrants as in Fig. 3.8)

Shifts in dates of maximum streamflow are the most significant of any index, and become increasingly significant as the cutoff year is changed from 1970, to 1975, to 1980 (Figure 4.3). While shifts are statistically significant in any quadrant, with the largest p-value at only 0.003, the mechanisms which seem to alter the timing of runoff likely vary greatly from region to region. In the southeastern states, many of which are fed by runoff from the southern Appalachian Mountains, runoff peaks have moved earlier in the season, most likely due to higher temperatures yielding earlier snow-melt. Evidence of earlier melting causing earlier peak streamflows can also be observed in certain mountainous catchments of New England and those in the Pacific Mountain ranges of the Sierra Nevadas and Northern Cascades.

However, in some Pacific catchments, peak runoff, once driven by the comparatively higher quantities of winter rainfall, has since become driven by melt as well, demonstrating how dramatically streamflow timing can shift in sensitive catchments (Littell et al, 2009). This is especially true in the Pacific Northwest where rainfall during winter is higher than any other location-time in the MOPEX dataset. In these cases, the shift from winter to spring appears in Figure 4.3 as a forward seasonal shift, though again increased melting appears to be the dominant process, as streamflow is increasing during a time period in which potential evapotranspiration rates approach their apex while rainfall rates are approaching their nadir.

In contrast, Figure 4.3 shows that runoff timing in the upper Midwest has shifted later in the season, often by nearly two months. This requires a different explanation, as the comparatively flat topography does not support mountain-top snowcaps. In this database's early years, the Midwestern streamflow peaks occurred during early spring as the frozen ground began to thaw. However, in recent decades, warmer climates may have prevented groundwater from freezing without interruption during winter months. As a result, the mechanism for maximum runoff is no longer the thawing ground but rather the timing of precipitation. In more recent years, maximum runoff in the Midwest arrives at roughly the same time as maximum precipitation. A shifting hydroclimatic regime, in these catchments, appears to have shifted not only the timing of maximum streamflow but the mechanism that drives it, from thawing to rainfall.

In the southern Midwest toward the Gulf of Mexico, although rainfall has generally occurred later in the season as previously discussed, runoff peaks have become earlier in the season in most cases, potentially due to earlier melting of snow packs in the Rocky Mountains that ultimately make their way into the lower Mississippi. Finally, in the northeast, many runoff

peaks have moved later in the season, again perhaps becoming more rainfall-driven (where the peak is later) than melt-driven (which occurs earlier).

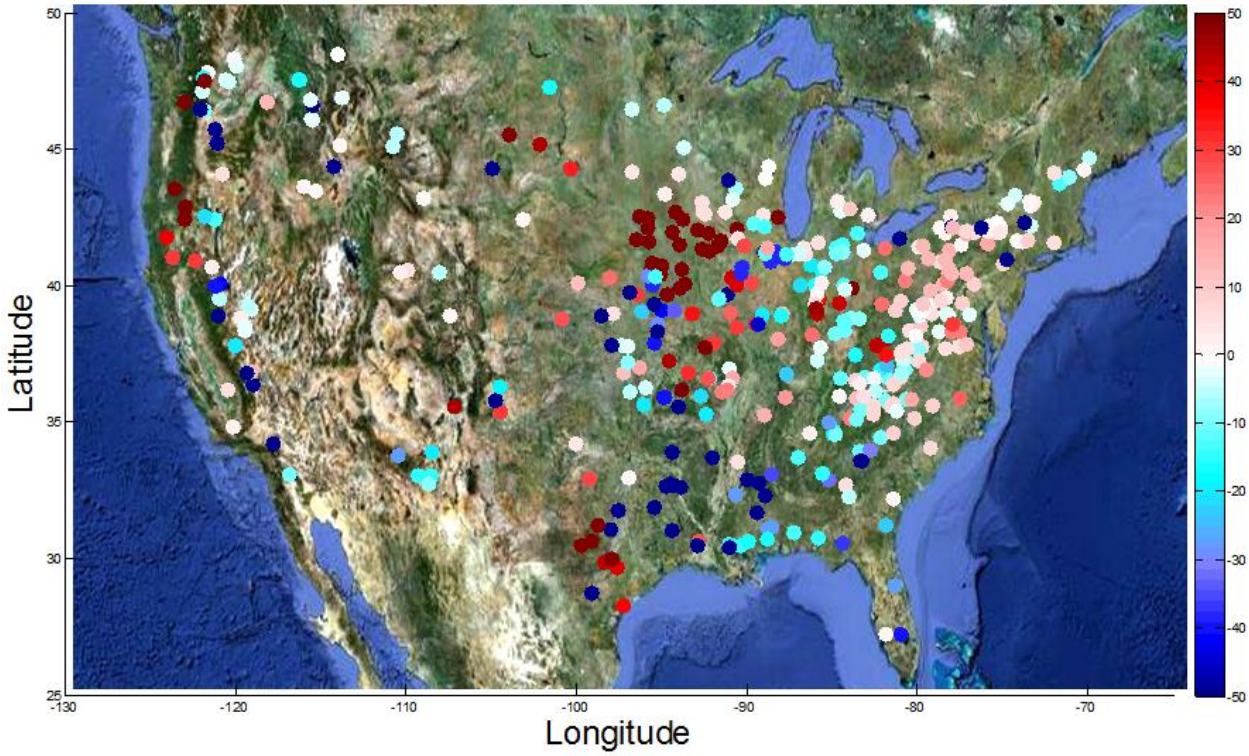


Figure 4.3 – Runoff Peak Date Shifts (in days), Pre/Post 1980
Color scale ranges from +50 (shifted forward, March to April, e.g.) to -50 (shifted backwards in time, April to March, e.g.)

4.3.4 Shifts in Classification: The Continental Story

In terms of the ultimate hydroclimatic classifications presented in Figure 1, catchments assume a different classification using only data before or after the threshold years. The most significant changes can be viewed via the four quadrants (Figure 3.8). If the data are divided at 1970, the most notable changes are the dark blue region's (catchments with low seasonality and rainfall peaks before June 1st) expansion north and east, replacing the light-blue region (catchments with low seasonality and rainfall peaks after June 1st). Elsewhere, the yellow region

(higher seasonality, non-arid) gives way near the southern Mississippi, becoming like its less seasonal eastern counterparts.

The same analytical procedure was performed using the two datasets formed by a division on 1980 rather than 1970. While the same qualitative results are observed, the extent is magnified tremendously when the cutoff year is advanced by a decade, as shown in Figure 4.4a and 4.4b. After 1980, the dark-blue region (catchments with low seasonality and rainfall peaks before June 1st) has covered a significant majority of the eastern half of the continent and is beginning to shift into many catchments in New England. The southern Midwest has seen sufficient loss of seasonality as to resemble what was once a region nearer the Great Lakes, although admittedly a much warmer version thereof. This is evidenced by the light and dark blue region nearly reaching the red region above the gulf in Figure 4.4b, whereas in 4.4a, a band of yellow still separates these two regions. These observations suggest that in the future, the northeastern catchments of the northern Appalachians, the catchments south of the Great Lakes, and those in New England will likely resemble the current behavior of the southern Appalachians and southeastern catchments. The southern Midwest will become increasingly less seasonal, resembling the eastern Midwest of years past (albeit warmer) if the current trend continues.

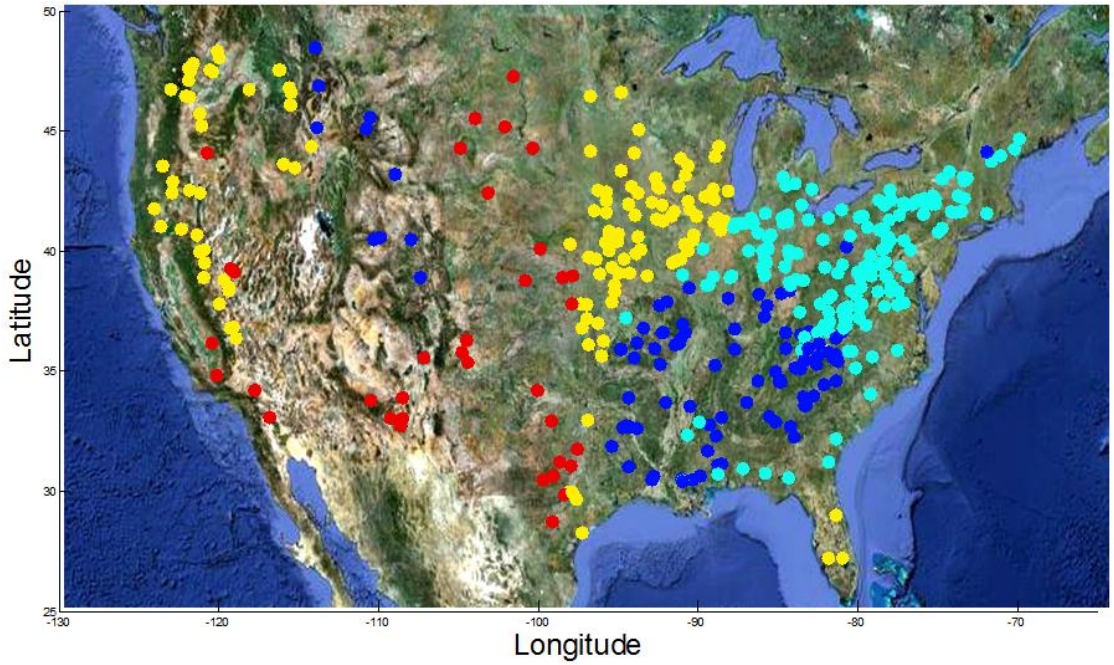


Figure 4.4a – The four quadrants from Coopersmith et al. (2012) Pre-1980
Low seasonality and earlier precip. peak (dark blue), low seasonality and later precip. peak (light blue), higher seasonality and non-arid (yellow), higher seasonality and arid (red).

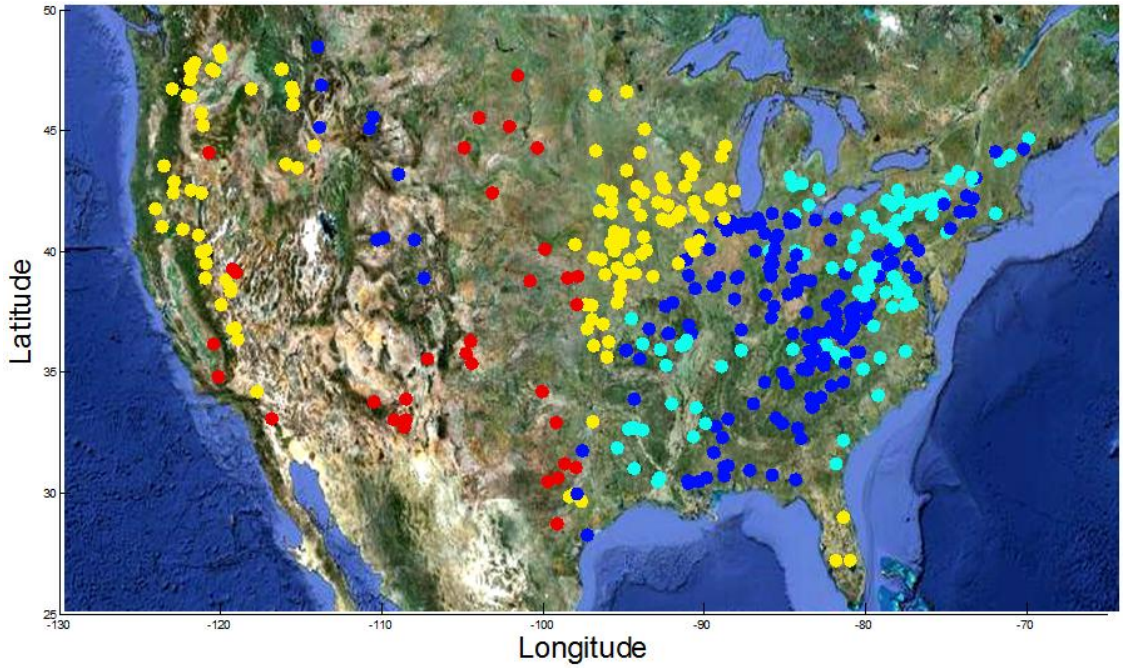


Figure 4.4b – The four quadrants, Post-1980
Low seasonality and earlier precip. peak (dark blue), low seasonality and later precip. peak (light blue), higher seasonality and non-arid (yellow), higher seasonality and arid (red).

4.4 Discussion: The Regional Stories

This section focuses on a few specific regions, either classes or groups of classes, whose regime curves with respect to rainfall and runoff display notable changes before and after the threshold years. In addition to an analysis of the regime curves themselves, changes in variability are explored through the maximum rainfall and runoff, minimum runoff (minimum rainfall is always 0), and standard deviation of rainfall/runoff quantities. All of these curves are smoothed by a 30-day moving average to avoid observations wholly determined by one storm or one dry-spell. This will allow confirmation of the findings of the previous sections and insights into specific changes in hydroclimatology that might influence decision making.

4.4.1 The Midwest: Change in Runoff Mechanism

Figure 4.5 (upper-left) displays the average hydrologic regime curve for the catchments classified as "Midwestern" (classes ITC, ISQJ, and ISCJ). The blue and redlines illustrate average daily precipitation before and after 1980, respectively. The blue-dashed and red-dashed lines represent average daily runoff before and after 1980. While the quantity of precipitation has increased slightly before and after these thresholds, the general shape is unchanged, reflecting a slightly warmer, wetter climate. However the runoff peak, as suggested in previous sections, has moved from early spring to early summer. This is consistent with the hypothesis that runoff has become rainfall-driven rather than the result of thawing as had been the case previously. Moreover, the overall quantity of runoff has increased by roughly 34% after 1980, despite the fact that precipitation has only increased by 8%. The runoff quantities after August (blue-dashed and red-dashed lines, after day 240) remain low, which supports the hypothesis that the earlier shifts are a result of changes in rainfall and thawing rather than land use changes.

Interestingly, while the average quantity of rainfall has increased throughout the year after 1980 (Figure 4.5, upper-left), the maximum rainfall by date (Figure 4.5, lower-left) was actually much larger before 1980. This suggests these less-frequent, larger rainfall events have become less common, but smaller storms are occurring more frequently. This is consistent with the observation that while total runoff has increased after 1980 (Figure 4.5, upper-left), maximum runoff values have not changed (Figure 4.5, lower-left). Moreover, this new pattern of more frequent, but less intense, rainfall has raised the minimum values for runoff after 1980 (Figure 4.5, lower-right).

In terms of decision-making, these shifts mean that the growing season is likely to begin sooner, as the ground will thaw much earlier than in years past. Furthermore, class ISQJ (Figure 1) has become more common. Before 1980, 69 catchments were classified as ITC or ISQJ. After 1980, 73 catchments fell into one of these two headings. Furthermore, 16 catchments originally classified as ITC switched to ISQJ after 1980. The reverse transition (ISQJ to ITC) occurred only twice. These shifts imply that many current ITC catchments may behave in the future as ISQJ catchments do today.

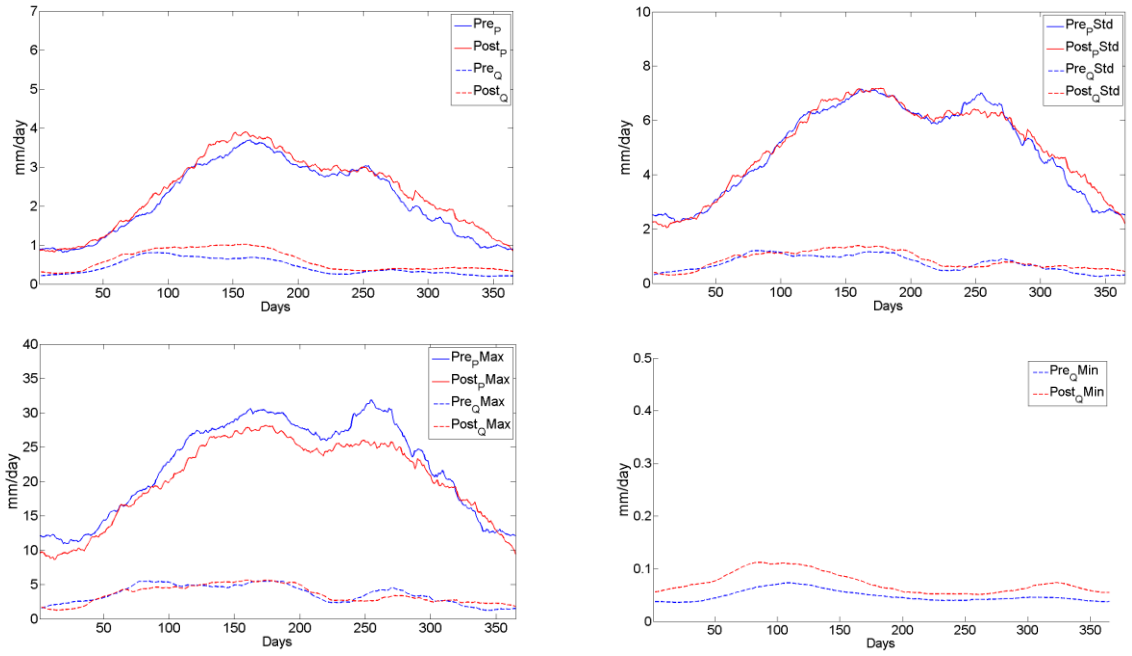


Figure 4.5 – Midwest, Before and After 1980
Upper-Left: Annual rainfall/runoff regimes, 30-day moving average
Upper-Right: Standard deviation of rainfall/runoff by date, 30-day moving average
Lower-Left: Maximum rainfall/runoff by date, 30-day moving average
Lower-Right: Minimum runoff by date, 30-day moving average

4.4.2 The Pacific Northwest: Smaller Snow Packs, Unchanged Winter Runoff

In Figure 4.6 (upper-left), the average hydrologic regime curves are presented for catchments in the Pacific Northwest (classes XHD, XVM, XTM, ISCB, and IHM). As would be expected, the runoff peak is decidedly lower after 1980 (red-dashed curve), as the snow packs have decreased steadily in size as a result of warmer temperatures (Nolin, 2012). The dates of maximum rainfall during winter (blue and red lines, days 330-65) are shifted later in the season and the rainfall peaks are diminished. However, winter runoff is roughly consistent before and after 1980. In terms of decision making, as winter rainfall peaks lessen, snowcaps shrink, and runoff quantities decrease, these catchments will likely begin to demonstrate hydrologic regimes closer to their southern neighbors (XSMB and XSC, see Figure 1).

Though diminished snow packs yield a lower quantity of runoff after 1980 (unsurprisingly), the maximum and minimum flows (Figure 4.6, lower-left and lower-right) during the fall and winter months (where snow melt is not occurring) are essentially unchanged. Despite the warming of Pacific waters as seen in the Pacific Decadal Oscillation, which shows significantly higher temperatures after 1980 than before, the variability of rainfall and runoff are relatively unchanged year-round (Figure 4.6, upper-right). It seems as though this warming of the proximal Pacific waters manifests in terms of diminished snow packs and earlier, lesser minimum flows (Figure 4.6, upper-left and lower-right respectively), but does not affect rainfall or runoff variability notably. These lower minimum flows may impact local industry built in a region where difficulties from limited available streamflow had been rare.

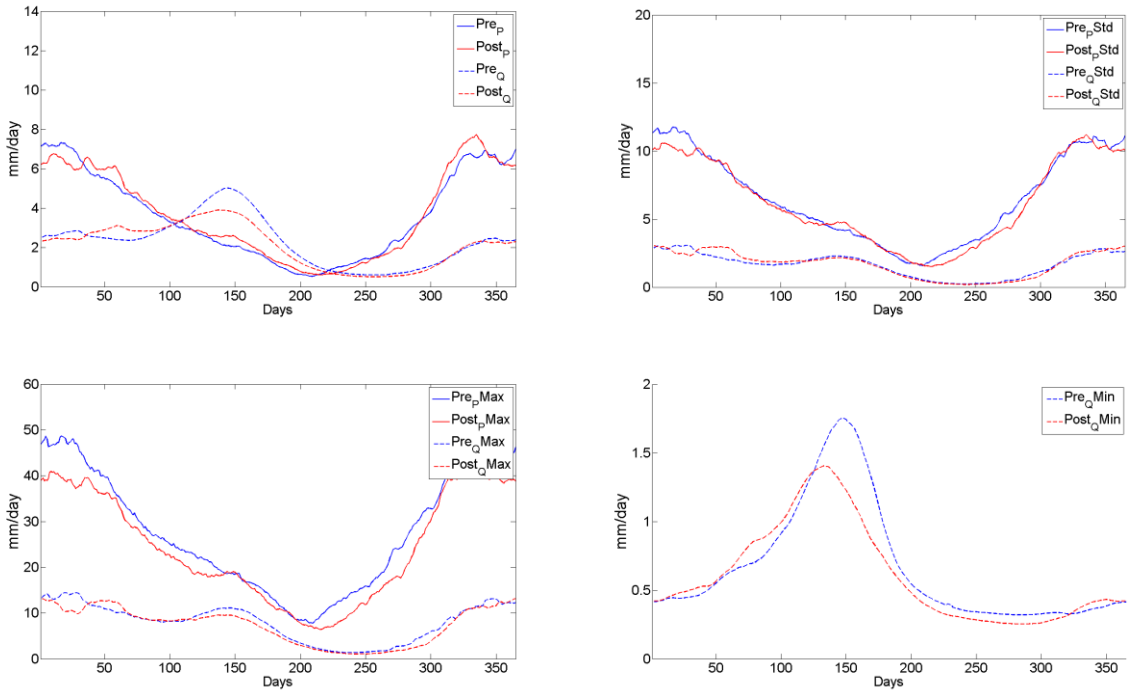


Figure 4.6 – Pacific Northwest, Before and After 1980
Upper-Left: Annual rainfall/runoff regimes, 30-day moving average
Upper-Right: Standard deviation of rainfall/runoff by date, 30-day moving average
Lower-Left: Maximum rainfall/runoff by date, 30-day moving average
Lower-Right: Minimum runoff by date, 30-day moving average

4.4.3 The Pacific Southwest: A New Spring Runoff Peak, Fewer Low Flows

Like their neighbors to the north, the Pacific Southwest (classes XACJ, XADB, XSC, and XSMB in California and Nevada) is characterized by maximum rainfall during the winter months. In Figure 4.7 (upper-left), it is clear that not only is the peak rainfall roughly one month later after 1980 than it had been before, but the peak is notably larger as well. This more intense rainfall during the late-winter/early-spring has also added a small runoff peak during early March (red-dashed line, days 40-65) that was non-existent before 1980. Moreover, the runoff has increased overall, spread out over a period from early-spring to mid-summer, although these catchments are still decidedly arid. It is possible that this increase in spring runoff could signal an earlier opportunity to plant for the upcoming growing season, though it is also possible that the increase in early-spring rainfall (red-dashed line, days 40-100) could erode soils that have rarely experienced this type of rainfall intensity (the blue line never reaches the peak of the red line).

Again, though the PDO has warmed the nearby coastal waters after 1980, the variability of rainfall and runoff (Figure 4.7, upper-right) has remained relatively consistent, with changes that seem to mirror the changes in the magnitude of rainfall and runoff. However, the minimum values of runoff (Figure 4.7, lower-right) are much higher throughout the year after 1980 and peak much sooner, despite very similar quantities of rainfall. This suggests that the small increase in rainfall and its later arrival after 1980 precludes the extreme low flows seen in previous years. The diminished risk of low flows may result in more favorable crop insurance prices and more stable crop yields in additional to fewer shortages domestically and industrially.

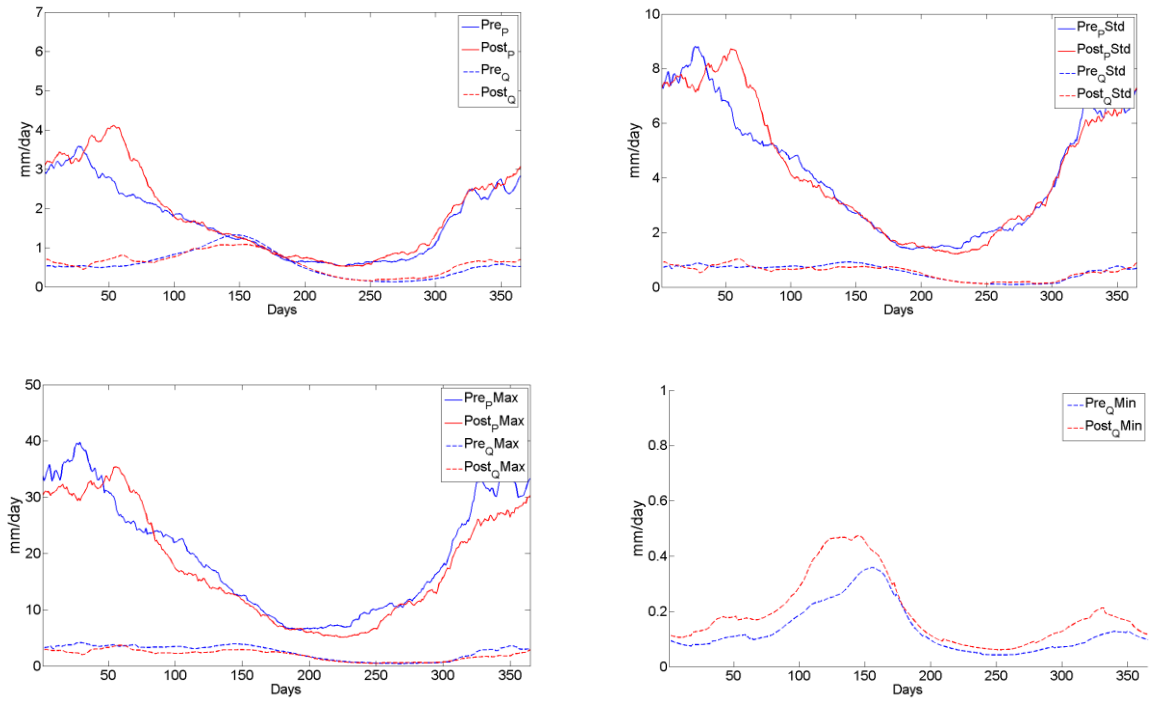


Figure 4.7 – Pacific Southwest, Before and After 1980
Upper-Left: Annual rainfall/runoff regimes, 30-day moving average
Upper-Right: Standard deviation of rainfall/runoff by date, 30-day moving average
Lower-Left: Maximum rainfall/runoff by date, 30-day moving average
Lower-Right: Minimum runoff by date, 30-day moving average

4.4.4 The Rocky Mountains: Smaller Snow Packs, Steadier Rainfall

Similar to the Pacific Northwest, Figure 4.8 (upper-left), an average regime curve for classes LPQ, LBMH and LBMS, illustrates the same diminished runoff peak in the Rocky Mountains during spring, albeit to a lesser extent, and later maximum rainfall. These effects may be less pronounced due to greater geographic distance from the coast, and thus a diminished effect of the warmer waters described by the PDO. Interestingly, although the runoff remains minimal during winter months before and after 1980 (blue-dashed and red-dashed lines, days 300-75, or November to mid-March), the rainfall throughout the year has increased significantly

in the years after 1980. This may indicate opportunities to grow crops that have lacked the necessary rainfall in previous decades.

Despite these slightly diminished snow packs, maximum and minimum runoff values (Figure 4.8, lower-left and lower-right) are essentially unchanged. However, maximum daily rainfall values are higher throughout the year before 1980, while total rainfall is higher thereafter. Much like the observations in the Midwest, the hydroclimate seems to favor more-frequent, but less-intense storms after 1980. This steadier rainfall pattern is corroborated by the lower variability of rainfall (Figure 4.8, upper-right) after 1980. More frequent and less intense storms may result in more predictable crop yields in certain locations, but diminished flows in downstream locations reliant on runoff events from large mountain storms may lead to scarcity concerns for industrial and agricultural decision-makers.

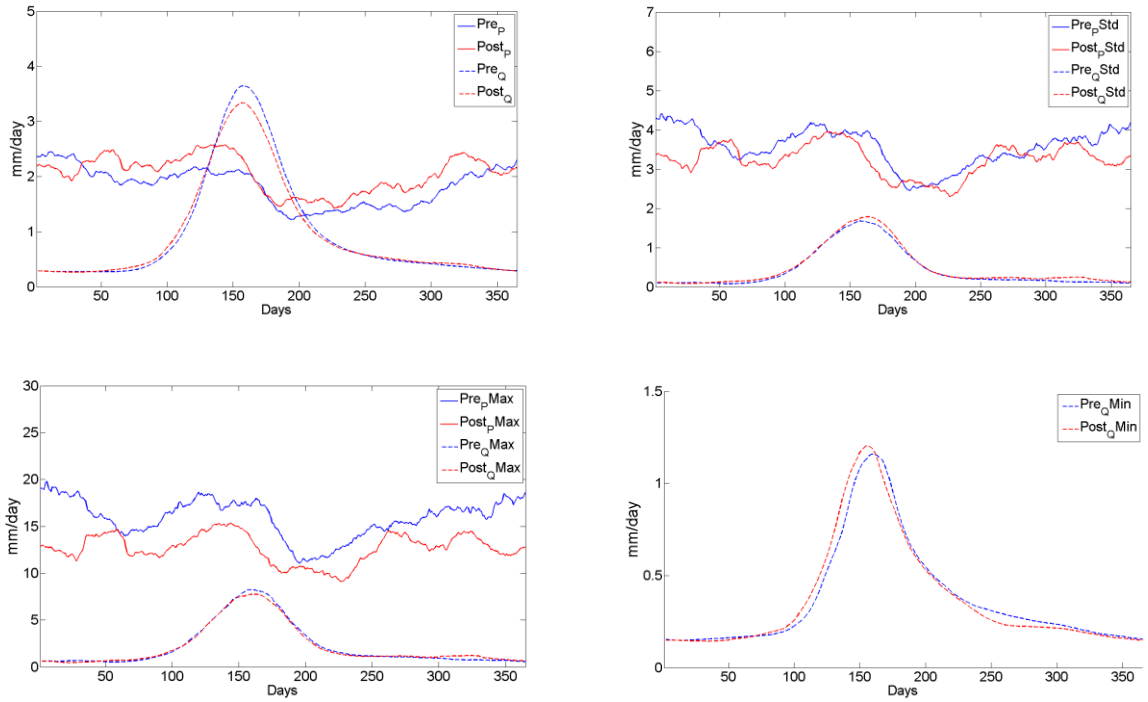


Figure 4.8 – Rocky Mountains, Before and After 1980
Upper-Left: Annual rainfall/runoff regimes, 30-day moving average
Upper-Right: Standard deviation of rainfall/runoff by date, 30-day moving average
Lower-Left: Maximum rainfall/runoff by date, 30-day moving average
Lower-Right: Minimum runoff by date, 30-day moving average

4.4.5 The Southern Midwest: Class IAF, Later Arrival of Intense Rainfall

In the Southern Midwest catchments (in Texas), whose average regime curve is illustrated in Figure 4.9 (upper-left), runoff is virtually nonexistent throughout the year. However, the patterns of annual precipitation display shifts in two peaks of the distribution. The first peak, occurring in late spring like most Midwestern catchments, is driven by seasonal rainfall consistent with many catchments of similar longitude. However the second peak, occurring in late-summer or early-fall, is driven by tropical storms and hurricanes, as evidenced by catchments of a similar longitude located farther north (away from the Gulf) that do not display this peak. With increased temperatures, these storms can occur later in the season,

extending the timeframe of the second peak as shown in Figure 4.9 (upper-left). While this may be beneficial in extending the length of the growing season, the cost of crop insurance is likely to increase, as the risk of extreme precipitation becomes more severe.

While the variability of rainfall (Figure 4.9, upper-right) shares a bi-modal distribution that seems similar to the rainfall signature, overall variability has increased since 1980, especially during the winter months when previously, rainfall quantities had been rather consistent. However, the maximum rainfall quantities (Figure 4.9, lower-left) have shifted forward in time but not increased, and runoff values are still very low due to the aridity of the location. Thus, the new-found variability does not reflect more intense storm events, but rather, the propensity to receive them during months in which they were previously rare. Decision-makers must therefore consider the possibility of extreme events during a wider range of months. New construction will need to reflect the longer-period of extreme weather risks.

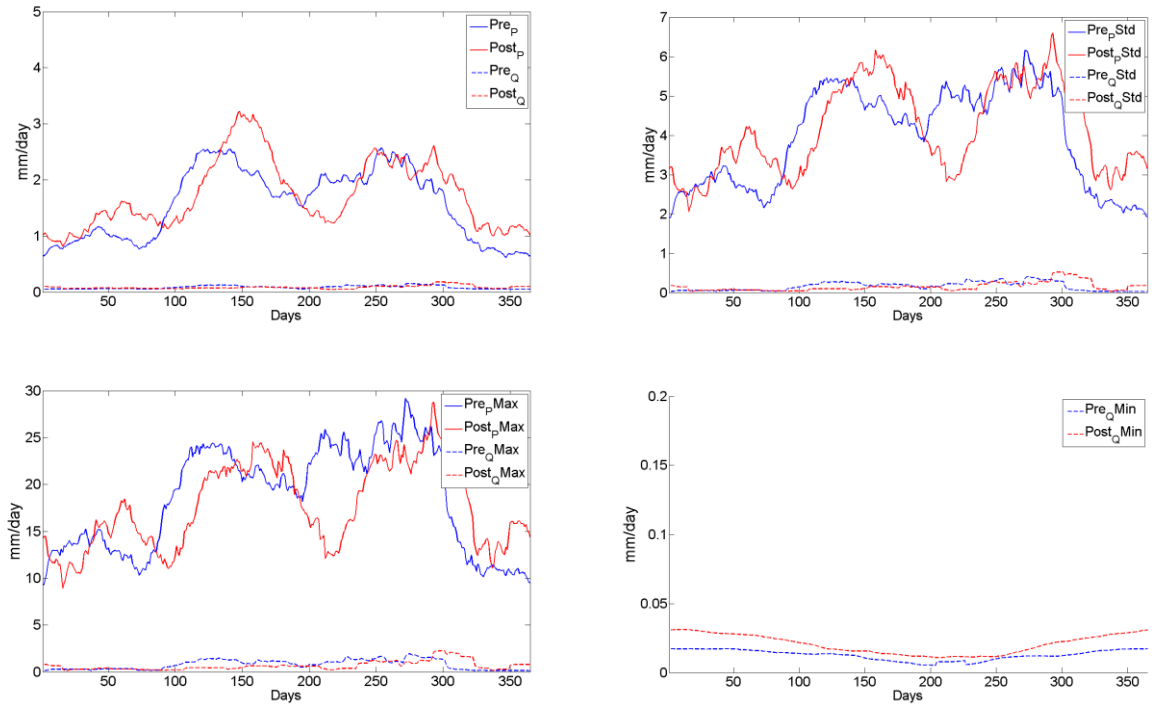


Figure 4.9 – Texas (Southern Midwest), Before and After 1980
Upper-Left: Annual rainfall/runoff regimes, 30-day moving average
Upper-Right: Standard deviation of rainfall/runoff by date, 30-day moving average
Lower-Left: Maximum rainfall/runoff by date, 30-day moving average
Lower-Right: Minimum runoff by date, 30-day moving average

4.4.6 The Southern Appalachians: Systemic Drying

Southern Appalachian catchments (Alabama, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee and Virginia, along with two similar northern catchments from Pennsylvania and Connecticut), which are in class LWC, exhibit a cluster of hydroclimatic behavior that, after 1980, represents a majority of all catchments in the eastern United States. For these catchments, the changes appear in terms of slightly earlier rainfall and earlier, diminished runoff (Figure 4.10, upper-left). These shifts, though lessened by the more constant rainfall typical of this region, tell the story of a climate becoming warmer, drier, and characterized by earlier melting of any mountain snow that feeds the streams. Other

proximal classes (LPM and LPC to the northwest and LJ to the northeast) all demonstrate similar properties. These catchments may offer a preview of the northeastern regions of the United States under the effects of hydroclimatic changes over the next several decades. While the increase in temperature may create new arable farmland in flatter regions of New England, many of the cold-weather crops (fruits, tree nuts, berries, etc) currently grown in the northeastern United States may struggle to flourish under a warmer climate.

The general sense from this analysis is a climate becoming drier. While the Pacific Decadal Oscillation suggests warmer ocean temperatures during the post-1980 data, the Atlantic Multidecadal Oscillation (AMO) places 1980 firmly in the middle of a trough in terms of temperatures. Thus, the observations seen on the east coast are not solely the byproduct of cyclical changes in ocean temperatures. Maximum rainfall values (Figure 4.10, lower-left) are lower at every date after 1980, and unsurprisingly as a result, the maximum and minimum runoff values (Figure 4.10, lower-left and lower-right) have decreased as well. Interestingly, the variability of rainfall/runoff (Figure 4.10, upper-right) is essentially unchanged. This suggests the results are not simply the manifestation of a large drought or two, but rather, a systemic decrease in the amount of water processed by these catchments. In terms of decision-making, scarcity may become an issue where it was not in the past, placing livestock and local industry and risk, and crop insurance contracts may become much more expensive.

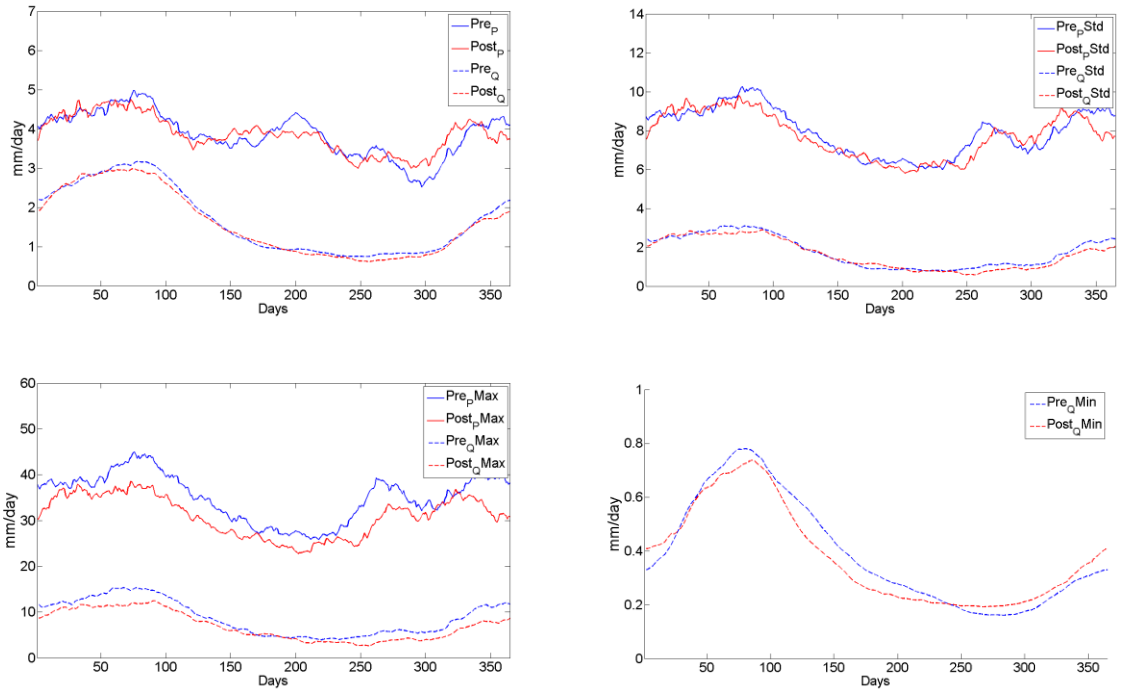


Figure 4.10 – Class LWC (Southeast), Before and After 1980
Upper-Left: Annual rainfall/runoff regimes, 30-day moving average
Upper-Right: Standard deviation of rainfall/runoff by date, 30-day moving average
Lower-Left: Maximum rainfall/runoff by date, 30-day moving average
Lower-Right: Minimum runoff by date, 30-day moving average

4.4.7 Florida (Class ITF): More Variable Runoff

Figure 3.13 illustrates that Floridian catchments are different from other catchments. This is confirmed in Figure 3.8, as the light-blue/dark-blue region consumes the entire eastern half of the continent, excluding Florida. In Figure 4.11 (upper-left), we observe fairly similar rainfall signatures (slightly more rain is received after 1980). However, the quantity of runoff observed is lower at every date.

This seems well-explained by the maximum rainfall/runoff signatures (Figure 4.11, lower-left), as the maximum rainfall values obtained are lower at essentially every date, yielding lower peak runoff, especially during the summer and fall. This suggests that runoff, previously

driven by unusually large storms, now has been diminished by slightly less-intense storms (and potentially, drier soils capable of greater storage). Observations of seasonal variability of runoff (Figure 4.11, upper-left) illustrate that after 1980, the variability of runoff has increased, perhaps due to decreased hydrologic residence times, especially during the first few months of the year. This shift may be caused by rainfall exiting the watershed more rapidly, causing higher variability than the gradual release of runoff. Agricultural and industrial decision-makers will begin to grapple with the possibility of limited streamflow.

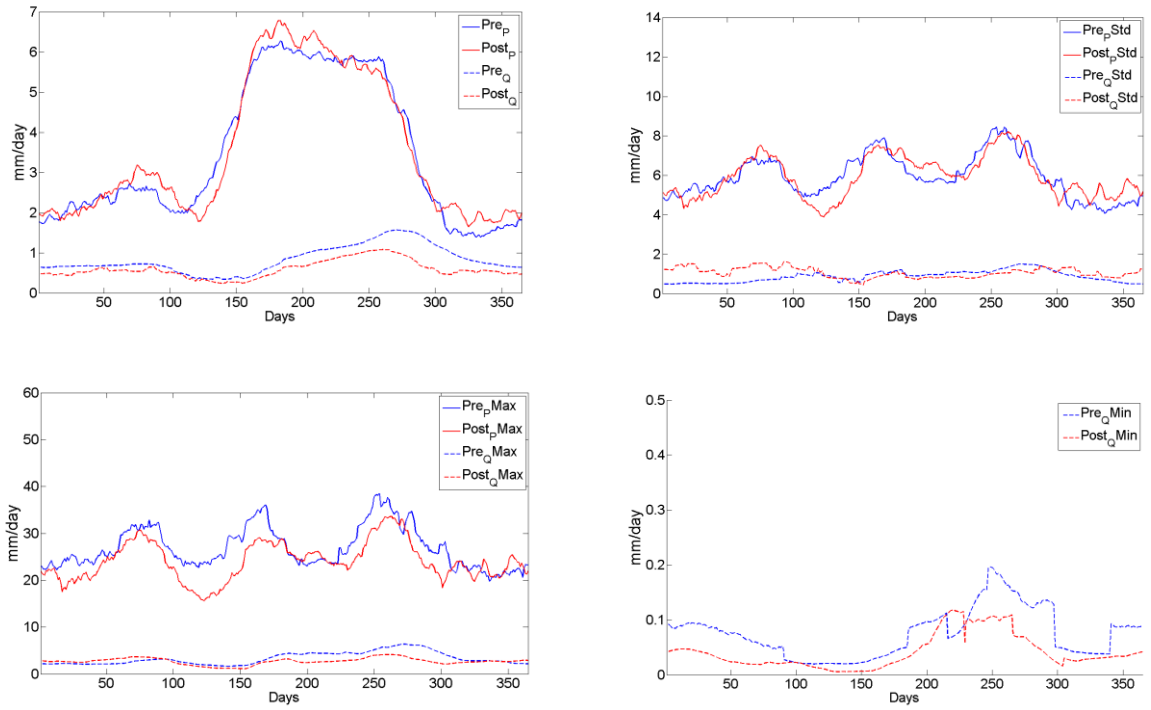


Figure 4.11 – Class ITF (Florida) Before and After 1980
Upper-Left: Annual rainfall/runoff regimes, 30-day moving average
Upper-Right: Standard deviation of rainfall/runoff by date, 30-day moving average
Lower-Left: Maximum rainfall/runoff by date, 30-day moving average
Lower-Right: Minimum runoff by date, 30-day moving average

4.4.8 The Northeast: More Stable Runoff

Despite the AMO's illustrations that the pre-1980 and post-1980 samples are similar in terms of their average ocean temperature (1980 is the middle of a trough in the AMO), changes in hydrologic regime patterns nonetheless are occurring in these catchments. After 1980, slight increases in total precipitation and its seasonality (Figure 4.12, upper-left) do not alter runoff patterns dramatically. Moreover, the variability with respect to rainfall is similar before and after 1980 (Figure 4.12, upper-right). However, the maximum and minimum regimes with respect to daily rainfall/runoff (Figure 4.12, lower-left and lower-right) tell a different story. The maximum rainfall has decreased at every date. The maximum runoff has decreased after 1980 while the minimum runoff has increased. This stabilization of maximum and minimum runoff values in more recent years seems to suggest longer hydrological residence times, rendering catchments less susceptible to the impacts of brief, intense storms. This shift may result from rainfall exiting the system more slowly, which results in lower runoff variability. These decreases in variability are likely favorable results for industrial decision-makers, now able to be more confident in the quantity of streamflow they will receive.

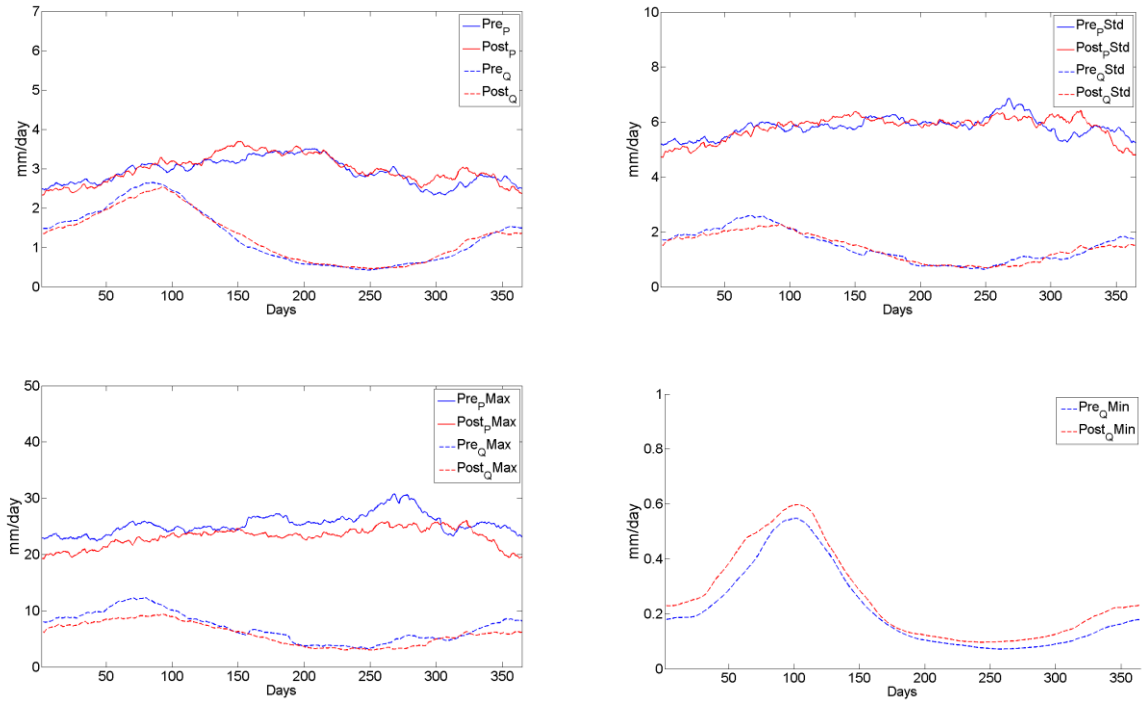


Figure 4.12 – Class LJ (Northeast) Before and After 1980
Upper-Left: Annual rainfall/runoff regimes, 30-day moving average
Upper-Right: Standard deviation of rainfall/runoff by date, 30-day moving average
Lower-Left: Maximum rainfall/runoff by date, 30-day moving average
Lower-Right: Minimum runoff by date, 30-day moving average

4.5 Conclusions and Opportunities for Further Research

Hydroclimatic changes have occurred throughout the United States, but their impacts differ by region. In the Midwest, where runoff once peaked during spring thawing, peak runoff now aligns with peak rainfall during summer months. In the Pacific Northwest, shrinking snow-packs decrease runoff during springtime, but winter runoff remains unchanged. In the Pacific Southwest, a new runoff peak appears during the spring and low flows have become less frequent. In the Rocky Mountains, snowpacks have been diminished, but rainfall has become less variable. In the southern Midwest, intense rainfall events are observed later in the season than in previous decades. In the Southern Appalachians, the quantity of water processed by

catchments has decreased substantially. In Florida, runoff has become more variable while in the Northeast, runoff has become more stable. Each region tells a different story. Continentally, the behavior that characterized the southeastern quarter of the United States before 1980 has spread northward and westward thereafter. In many catchments, a preview of conditions in years to come may already exist elsewhere.

These regional and continental stories are important for decision-makers, whether in terms of water scarcity, the growing cost of crop insurance, changes in the length and timing of growing seasons, the need for irrigation, changes in the selection of optimal crops, and the consequences for risk management in regions likely to receive more frequent instances of severe droughts and storms. To observe the potential impacts of subsequent hydroclimatic shifts, one can look to catchments that have already experienced such shifts and note the responses in hydroclimatic and managerial terms. This type of prognostication becomes computationally, logistically, and administratively unmanageable if each plot, or even each catchment, requires individual analysis. The notion of hydroclimatic classification and the subsequent merging of these classes allows for more robust qualitative insights at the regional and national scale to emerge.

While this work does not forecast changes in rainfall and runoff for the decades to come, predictive models could be used to analyze these regional stories in more detail in future research.

Ultimately, this analysis displays changes in classification, rainfall and runoff regimes, variability of these regimes, and daily maximum or minimum rainfall/runoff events. Further research is needed to analyze individual regions in more detail, contextualizing these findings with datasets that speak to localized, specific issues (e.g., more frequent, less intense storms,

lengthened growing seasons, more arid climates, or more or less variable runoff). This analysis is not intended as the final word on regional hydroclimatic changes, but rather, the beginning of a more focused discussion.

Having developed hydroclimatic classifications in Chapter III and an understanding of their potential changes in chapter IV, we begin to apply those classifications as a means of regionalizing and generalizing soil moisture models in Chapter V.

CHAPTER 5

ADVANCES IN SOIL MOISTURE MODELING: A GENERALIZATION OF THE DIAGNOSTIC SOIL MOISTURE EQUATION USING CLASSIFICATION

After developing a mechanism of clustering groups of hydroclimatic similarity at the catchment level in Chapter III and a mechanism to assess how these regions have changed throughout the past several decades in Chapter IV, this chapter presents a physically-based soil model from Pan et al (2003, 2012) that is subsequently improved by the application of machine learning techniques. The question becomes, once this model is calibrated successfully, what mechanism will enable the application of those calibrations elsewhere? More succinctly, "how will we know where parameters can and cannot be applied?" The answer will be found in the delineations of hydroclimatic classes from Chapter III. In demonstrating that a calibration at a single site can be reapplied elsewhere within the same hydroclimatic class, the physically-driven soil moisture model, previously only applicable at specific sites at which it has been calibrated, can be utilized anywhere. Moreover, by integrating these soil moisture estimates with LiDAR-driven elevation data, soil moisture estimates can be adjusted based upon local topography, enabling fine-scale estimates, even with only a single sensor, located elsewhere within the same hydroclimatic class.

As individual, qualitative assessments of wetness/dryness conditions will be unavailable at the majority of relevant locations, a separate, objective metric will be required. For the purposes of this work, soil moisture will represent a proxy measurement of wetness. This type of data is available at a variety of sites across the continental United States and, if modeled well at

these locations using public source of climate data, will facilitate appropriate decision-support from local land-managers using these models.

To address the shifting climate classes seen in Chapter IV, the analysis will not only address shifting climatic regimes within regional classes, but also inform forecasts of which locations are most likely to undergo changes in climate behavior that will ultimately require managerial action in terms of crop selection, use of irrigation, or when the first planting of the season ought to occur.

Machine learning models constructed from local, public climate data can facilitate decision-support, integrated with hydroclimatic classification modeling to address large-scale hydroclimatic trends in space and time. These classes will enable the cross application of calibrated soil moisture models, enabling decision-support wherever it is requested.

5.1 Introduction: The Need For High-Granularity Soil Moisture Estimates

Efforts to glean soil moisture values from remotely sensed data have permeated nearly forty years of scientific literature, beginning perhaps with microwave radiometers mounted on aircraft flying over an agricultural site in Arizona (Schmugge et al, 1974). These tests verified, at least in principle, the plausibility of measuring soil moisture from overhead. Springing forward 40 years, the European Space Agency's Soil Moisture and Ocean Salinity (SMOS) mission has deployed satellites to estimate soil moisture over large swaths of land, demonstrating strong performance against in situ sensory grids (Jackson et al, 2012). However, while these satellite-based sensors can photograph the entire globe, the spatial resolution of their measurements (several kilometers to tens of kilometers) often proves inadequate for the decision-support end-users ultimately require. Even when NASA's Soil Moisture Active-Passive (SMAP)

satellite launches in the fall of 2014, estimates at a resolution finer than 3km will be inaccessible without more sophisticated modeling work. To this end, soil moisture models such as those from Pan et al (2003, 2012) are designed to produce soil moisture estimates from precipitation data, enabling estimates of soil moisture in the absence of in situ soil moisture sensors.

The diagnostic soil moisture equation (Pan et al, 2012), requiring only a time series of precipitation data, a simple sinusoidal estimate of potential evapotranspiration, and three calibrated parameters (residual soil moisture, porosity, and conductivity/drainage), offers a theoretical opportunity to address this shortcoming of satellite estimates. This algorithm, in its elegance and simplicity, has chosen to omit certain hydrologic occurrences, most notably the impacts of snow cover and its subsequent melting. The model presumes that precipitation increases soil moisture upon arrival (unless the soil is saturated) and that the absence of precipitation gradually dries the soil (unless the soil is maximally dry). This assumption's logic notwithstanding, soil covered in snow does not become wetter as more snow falls, nor does it become drier when the sun and warmer temperature begin melting the snow. The calibrations from Pan et al (2012) took place at sites where snow virtually never occurs (New Mexico, Georgia, etc). The decision support services this work aspires to enable cannot be relegated to solely warm, dry locales. Thus, the calibration in this work is focused upon the growing season (April – October). Though one recognizes that the models developed will ultimately function optimally only seven to eight months per year, these months represent the time period in which detailed soil moisture estimates will prove beneficial to most applications. To wit, a snow-covered cornfield does not require advanced, real-time modeling to reveal this fact.

This diagnostic soil moisture equation is a simplified, lumped model that was chosen over other soil moisture models (HYDRUS-1D, Simunek et al, 1998, e.g.) for several reasons. First,

more complex soil moisture models, such as HYDRUS, require detailed knowledge of hydraulic soil parameters, information regarding root structures, soil temperature readings, ionic chemistry, CO_2 concentrations, solute transportation data, and detailed atmospheric/meteorological information. The diagnostic soil moisture equation requires only two simultaneous time series – precipitation and soil moisture. Second, its simplicity facilitates practical implementation of its code in a manner that can be easily manipulated and refit with new parameters (via genetic algorithm or via cross-application). To apply a more complex model at a different location from the one at which it is calibrated increases the probability of poor performance of one of its numerous inputs that may be inappropriate at the new location. Third, though the lumped nature of the diagnostic soil moisture equation does not address vertical variability of soil moisture, this shortcoming is addressed by separate calibration at two-inch and four-inch depths only.

This work hypothesizes that calibrated soil moisture parameters can be re-applied at hydroclimatically similar locations. This hypothesis is validated via a definition of hydroclimatic similarity that builds upon the classification system developed in Chapter III. However, it is important to recognize that in terms of predicting soil moisture using parameters calibrated elsewhere, while hydroclimate plays an important role in defining "similarity," so too do soil properties. While this feature is not the primary focus of this analysis, the impacts of soil texture and type are addressed within a subsequent discussion section.

Similar to Pan et al (2012), this work calibrates the diagnostic soil moisture equation via data from the U.S. Department of Agriculture's (USDA) Soil Climate Analysis Network (SCAN). This national array of soil moisture sensors (with co-located precipitation sensors) delivers hourly data at a variety of publically-accessible sites throughout the United States. Fifteen sensor locations with numerous years of high-quality, minimally-interrupted data were

selected for further analysis. These sites display notable hydrologic diversity, which will aid in demonstrating that the nationwide application of such soil moisture models using precipitation data represents a feasible goal.

5.2 Methodology

The approach has three steps. First, the diagnostic soil moisture model is calibrated (during the growing season) at locations with ample data. Second, the predictions at these locations are improved using machine learning techniques for error correction. Third, and most importantly, the classification system proposed by Coopersmith et al (2012) is used to generalize the parameters calibrated at each location – verifying whether the diagnostic soil moisture equation is calibrated successfully at a single site, those parameters enable its application at other sites characterized by the same hydro-climatic class. Each step is described in more detail in the sections below.

5.2.1 Calibration: Two-Layer Genetic Algorithms

Unlike the original diagnostic soil moisture calibrations, the objective of this work, in the broadest sense, lies in enabling decision-support for end users in real time. To this end, the daily model from Pan et al (2012) is modified to yield an hourly model within the same framework. SCAN sites provide hourly precipitation data and hourly soil moisture data with which to validate the results. Compared with Pan et al (2012), these models are calibrated over a higher number of years of data *and* twenty-four times the number of data points per day. Thus, while the original diagnostic soil moisture calibration deployed a Monte Carlo search of the parameter space to inverse the relevant non-linear equations, this is simply not practical without

supercomputing resources. To this end, genetic algorithms are deployed to explore the search space more efficiently.

Genetic algorithms, a subset of evolutionary algorithms, have been deployed since Barricelli's work in the 1950s and his subsequent development of an algorithm that simulated learning to play a simple game (1963). Such tools became more prevalent as optimization techniques following the work of Rechenberg and Schwefel in the 1970s (Rechenberg, 1973; Schwefel, 1975, e.g.). In recent years, these tools have become increasingly common in environmental and water resources applications, including the calibration of model parameters (e.g., Cheng et al, 2006; Singh et al, 2008; Zhang et al, 2009).

The standard form of the diagnostic soil moisture equation appears below:

$$\theta_{est} = \theta_{re} + (\phi_e - \theta_{re})\left(1 - e^{-c_4 \beta}\right) \qquad \text{Equation (5.1)}$$

Here θ_{est} represents the best estimate of soil moisture during a given hour. θ_{re} denotes residual soil moisture, the minimum quantity of moisture that is present regardless of the length of time without precipitation. ϕ_e, the soil's porosity, signifies the maximum possible soil moisture value, at which point the soil becomes saturated and cannot increase its moisture content. Finally, c_4 is a parameter related to conductivity and drainage properties, essentially defining the rate at which soil can dry. If c_4 assumes a value of zero, the soil is permanently at its residual soil moisture value, θ_{re} - a soil that dries infinitely rapidly. Conversely, as c_4 becomes large, the soil will permanently assume the value of its porosity, ϕ_e – a soil that dries infinitely slowly. The β term is presented in equation 5.2.

$$\beta = \sum_{i=2}^{i=n-1} \left[\frac{P_i}{\eta_i} \left(1 - e^{-\frac{\eta_i}{z}} \right) e^{-\sum_{j=1}^{j=i-1} \left(\frac{\eta_j}{z} \right)} \right] + \frac{P_1}{\eta_1} \left(1 - e^{-\frac{\eta_1}{z}} \right) \qquad \text{Equation (5.2)}$$

P_i denotes the quantity of rainfall during hour i (day in the original presentation in Pan et al). The soil depth at which an estimation occurs is given by z. This convolution summation has a temporal window of size n. Stated differently, n represents how far back in time the model must look for precipitation data. To wit, yesterday's rainfall impacts today's soil moisture, last week's rainfall is relevant, but less so, and rainfall from ten years ago is not relevant.

To choose the appropriate value for n, the value of β is calculated at each hour throughout the dataset – setting n to a very large value (2000 hours, denoted by M) initially. Next this "beta series" (where $n = M$) is correlated with a separate beta series, calculated where $n \ll M$. If the correlation between these two time series approaches unity, the smaller value of n is selected. Otherwise, n is increased incrementally until the correlation between the $n \ll M$ beta series and the $n = M$ beta series approaches unity. Finally, the η_i terms signify the estimated potential evapotranspiration / drainage loss at hour i of the calendar year. As this algorithm does not presume any more detailed knowledge of potential evapotranspiration/drainage behaviors, this "eta series" is modeled as a sinusoid (Pan et al, 2012) with period 8760 (the number of hours in a year).

The eta series is required to calculate the beta series (eq. 5.2), which is required to use the diagnostic soil moisture equation (eq. 5.1). Thus, before any other parameters are chosen, a sinusoidal estimate for the eta series must occur. To this end, a generalized sinusoidal form for η is given in equation 5.3:

$$\eta = \alpha \sin(i - \delta) + \gamma \qquad \text{Equation (5.3)}$$

In equation 5.3, α represents the sinusoid's amplitude, γ denotes the vertical shift, and δ signifies the necessary phase shift. These three parameters are fit via genetic algorithm such that the correlation between the beta series (using the eta series implied by α, γ, and δ) and the observed soil moisture series (θ_{obs}) is maximized. Once values for the eta series are established, the remaining three parameters of equation 5.1 (θ_{re}, ϕ_e, and c_4) are then fit by a second application of the genetic algorithm, this time minimizing the sum of squared errors between the estimated soil moisture series (θ_{est})and the observed values (θ_{obs}).

5.2.2 Error Correction

After the parameters of the diagnostic soil moisture equation (Eq. 5.1) have been calibrated, the hourly precipitation time series can be used to generate a soil moisture time series during the growing season months of interest. Naturally, there will be discrepancies between the observed soil moisture values (θ_{obs}) and the estimated values(θ_{est}), as expressed in equation 5.4:

$$\theta_{obs} = \theta_{est} + \varepsilon \qquad \text{Equation (5.4)}$$

where ε represents the error associated with any hour's soil moisture estimate. In attempting to correct these errors, the KNN algorithm is employed to predict ε using the characteristics from the training data. More specifically, the data are searched for the most similar matches in terms of time of day, day of year, θ_{est}, $\beta(n)$, and $\beta(M) - \beta(n)$. For example, if the model returns a prediction of $\theta_{est} = 0.35$ at 2:00pm during July when rainfall has been heavy recently but drier

over a longer period, KNN will search the training set for other estimates near 0.35 made on mid-summer afternoons where a similar recent rainfall pattern has been observed. Next, the algorithm averages the value of the error, ε, associated with those types of conditions, producing an estimated error, ε_{est}. Each validation estimate is then adjusted to be:

$$\theta_{est} + \varepsilon_{est} \qquad \text{Equation (5.5)}$$

This technique allows consistent model biases, such as underestimating wetter days and overestimating drier days, to be corrected.

Additionally, the diagnostic soil equation developed by Pan et al (2012) was designed to deliver daily soil moisture estimates. As a result, the eta series, used to describe the potential losses due to evapotranspiration and drainage is a sinusoidal function with a 365-day period. While this may describe seasonal changes in drainage and loss, the diurnal variations are overlooked. This would seem at odds with our intuition, which suggests that soil dries more rapidly during sunlit hours than after the sun has set. Using a machine learning algorithm for error correction indirectly models the diurnal cycle that the original diagnostic soil moisture equation does not consider. Consider a soil moisture estimate at 4pm, after soil has had a full day of sunlight (theoretically) to dry. As the diagnostic soil moisture equation only considers drainage and evapotranspiration losses on a daily basis, θ_{est} will be larger than θ_{obs}. Yet, because this type of mistake presumably occurred frequently throughout the training data, the algorithm will locate other 4pm estimates, each of which will be biased in the same direction, and our final soil moisture estimates will take this bias into account, improving the results.

5.2.3 Cross-Application and Cross-Validation

The two-part genetic algorithm calibration procedure detailed in the previous section allows six parameters to be determined for any site at which a precipitation time series and soil moisture time series are available. Generalizing the results and applying them in locations without soil moisture sensors requires additional information. The relevant scientific question is, given a calibrated soil moisture model at a specific location, where else can those parameters be applied? The hypothesis to be tested here is that the classification system by Coopersmith et al (2012) can be used to generalize the calibrated parameters for the diagnostic soil moisture equation by class. If two locations are assigned the same hydroclimatic classification, then the calibrated parameters from one SCAN sensor within that class will be assumed to perform well at another.

This hypothesis is tested at fifteen SCAN sensors for which soil moisture and precipitation data are available hourly for a period of several years. These sensors are located in diverse geographic locations and hydroclimatic classes in Iowa, North Carolina, Pennsylvania, New Mexico, Arkansas, Georgia, Virginia, South Carolina, Nebraska, Colorado, and Wyoming. The data at each of these locations were divided into training/validation sets and parameters were calibrated using training data only. Next, these parameters were employed on the validation sets at the locations for which they were calibrated. The subsequent R^2 values (proportion of variance in soil moisture explained by the machine-learning-enhanced diagnostic soil moisture equation) defined a baseline level of performance for that site.

The process of cross-validation is detailed below:

1. Consider two sites, x and y, chosen from the fifteen available calibrated locations.

2. Estimate the soil moisture values in the validation dataset of site y, using the parameters calibrated from the training dataset at site x.
3. Record the difference between the R^2 baseline value at site y (obtained using parameters calibrated at site y) and the performance obtained at site y using parameters calibrated at site x.
4. Repeat steps 1-3 for all 210 possible (x, y) pairs where $x \neq y$

   Note: (x, y) and (y, x) are not equivalent. One signifies the performance of parameters calibrated at site x making predictions at site y, the other signifies the performance of parameters calibrated at site y making predictions at site x.

At this point, three types of (x, y) pairs emerge. The first, when x and y fall within the same hydroclimatic class, should display limited losses in predictive power. The second, when x and y fall within a "similar" hydroclimatic class (two classes differing by a single division of the classification tree, developed in Chapter III) should display greater losses of predictive power. Finally, the third, when x and y fall in two unrelated classes, should display the largest loss of predictive power.

5.3 Results

5.3.1 Calibrations Before and After Machine Learning Error Correction

In this section, three sites in different hydroclimatic locations (New Mexico, Iowa, and Georgia) are examined in detail to illustrate how improvements from adding machine learning error models to the diagnostic soil moisture equation differ across sites. For these three sites, the

average quantity of additional variance explained by machine learning is 8.2%, with the full dataset of fifteen sites averaging an 8.3% improvement – these three datasets are representative in that respect. In each case, the six parameters required for the implementation of the diagnostic soil moisture equation are calibrated using training data from before 2010. Sensors with hourly precipitation and soil moisture time series data between 2004 and 2009 (inclusive) provide four to six years of training data (some sites are missing one or two years of data). Only days of the year where snow cover is unlikely are used to train the algorithm (from the 100th to 300th day of the year in all locations, for consistency). Validation data consist of days 100-300 for 2010 and 2011. At all three sites, the base model results are displayed alongside the results produced by deploying the machine learning algorithm (KNN) to remove bias and correct errors. In each image, the blue line represents the observed soil moisture readings, the red line represents the estimates generated by the diagnostic soil moisture equation, and the green line represents those predictions after the machine learning algorithm has removed biases and corrected errors. Soil moisture values (y-axis) are measured in percentage terms (0-100).

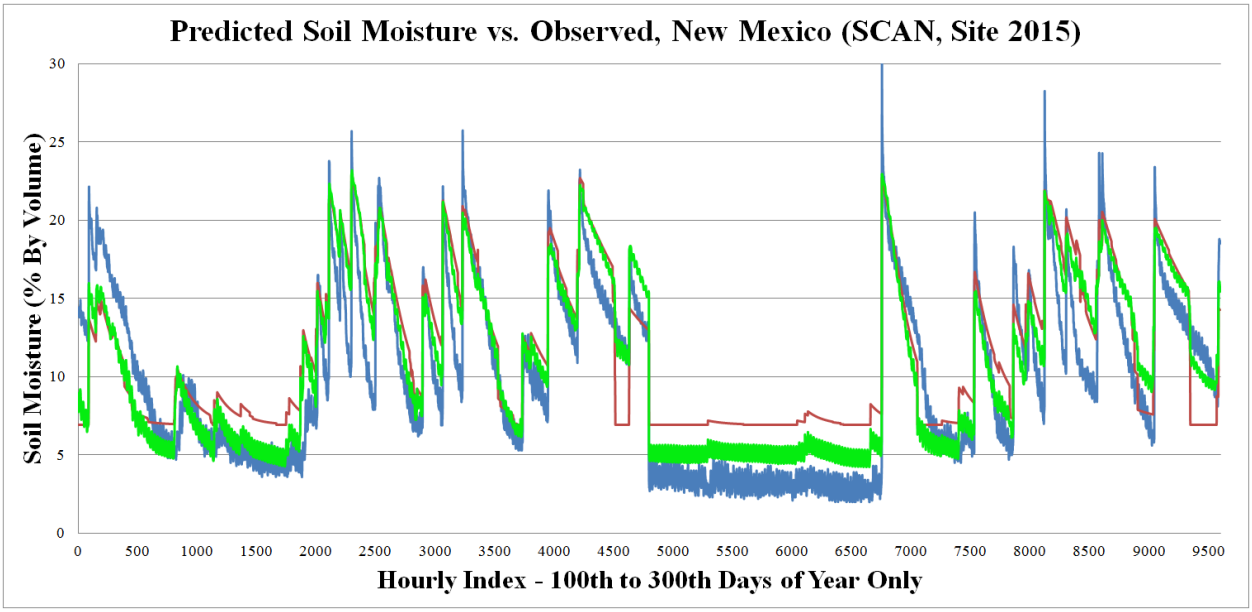


Figure 5.1, Soil Moisture Time Series, SCAN Site 2015, New Mexico (USA), Actual Soil Moisture (Blue Line), Diagnostic Soil Moisture Equation Estimate (Red Line), and Diagnostic Soil Moisture Equation with Machine Learning Error Correction (Green Line)

In figure 5.1, the diagnostic soil moisture equation is able to trace the general trend of the soil moisture time series ($\rho = 0.860$). However, during the middle of the time series, in which the observed soil moisture values fall below 5%, the benefits of machine learning error correction are most noteworthy. There are other hours scattered throughout the dataset where the green line (ML prediction) follows the blue line (observed values) much more nearly than the red line (diagnostic soil moisture equation). The green line ($\rho = 0.917$) not only improves upon the correlation value of Pearson's Rho (eq 5.6), but also displays marked improvement for those cases in which the diagnostic soil moisture equation produces significant errors.

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = \frac{E(X-\mu_x)E(Y-\mu_y)}{\sigma_x \sigma_y} \qquad \text{Equation (5.6)}$$

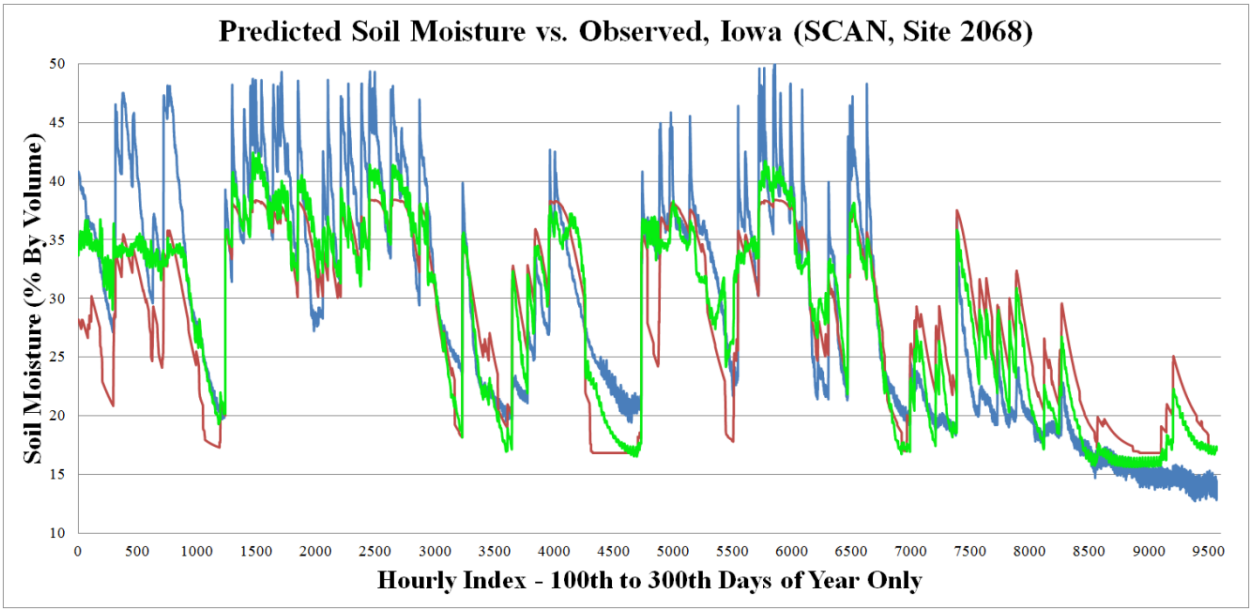


Figure 5.2, Soil Moisture Time Series, SCAN Site 2068, Iowa (USA), Actual Soil Moisture (Blue Line), Diagnostic Soil Moisture Equation Estimate (Red Line), and Diagnostic Soil Moisture Equation with Machine Learning Error Correction (Green Line)

During the validation period, considerable flooding occurred in Iowa (Figure 5.2). Flood events of this nature were not experienced during calibration. As a result, the porosity parameter is set at just above 41%. While this was appropriate for the training data (during which soil moisture did not exceed this level), extreme flooding events have caused moisture levels for which the diagnostic soil moisture equation was not properly calibrated. However the machine learning driven error correction improves the diagnostic soil moisture equation ($\rho = 0.846$) significantly ($\rho = 0.915$), even without remedying the errors due to exceedingly wet (flooded) soil. Underestimations due to floods, though detrimental in terms of numerical errors, are not necessarily a problem for decision support. If a model warns that the field is very wet and in reality, it is even wetter than predicted, the user has still been given adequate warning not to attempt activity at that site.

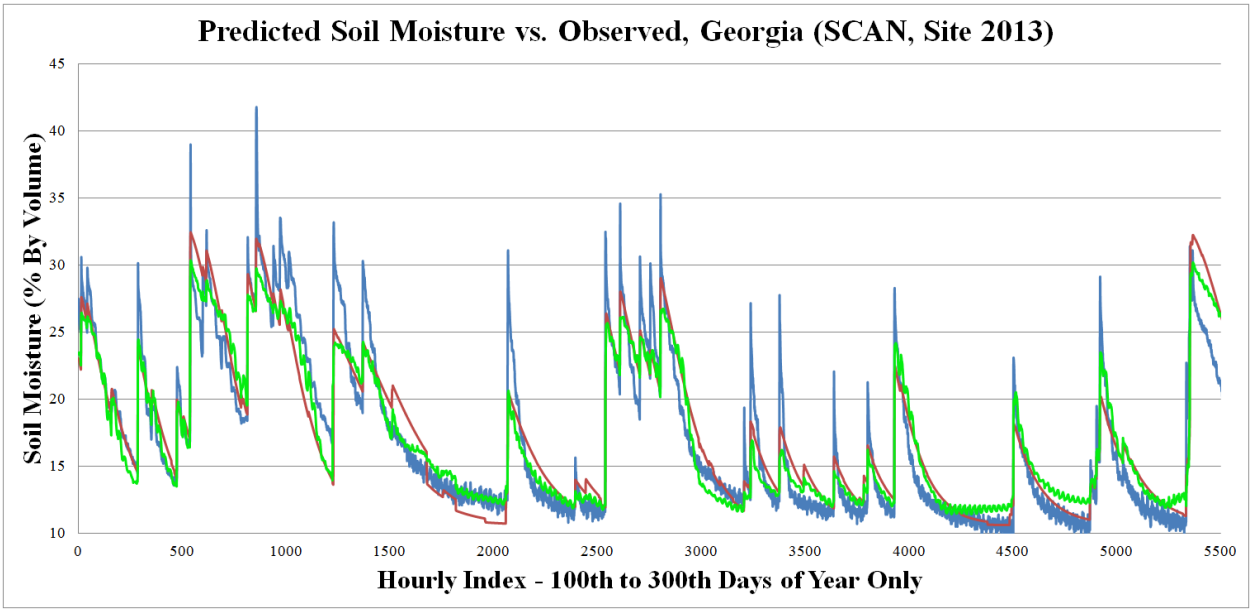


Figure 5.3, Soil Moisture Time Series, SCAN Site 2013, Georgia (USA), Actual Soil Moisture (Blue Line), Diagnostic Soil Moisture Equation Estimate (Red Line), and Diagnostic Soil Moisture Equation with Machine Learning Error Correction (Green Line)

In figure 5.3, a soil moisture series from Georgia is modeled by the diagnostic soil moisture equation. Even before adding any error correction, the equation performs well (ρ = 0.936) and the machine learning approach yields a smaller improvement (ρ = 0.941). It is worth noting that machine learning does not damage an already excellent performance, offering slight improvements when possible and essentially no correction when training data suggest the model has already performed adequately.

5.3.2 Cross-Validation Results: Qualitative Findings and Significance Testing

The fifteen SCAN sites calibrated for the purposes of this analysis yield $15^2 = 225$ possible (x, y) pairs. Fifteen of these 225 pairs occur when $x = y$, establishing the baseline level of performance for a given site (validation performed using the parameters calibrated at that same location). Of the 210 remaining (x, y) pairs, 120 of them consist of paired catchments in

which x and y are located in unrelated classes, 60 consist of paired catchments in which x and y are located in a "similar" class (different by a single split within the classification tree), and 30 consist of paired catchments in which x and y fall within the same hydroclimatic class (but x and y do not represent the same catchment). Figure 5.4 presents box plots illustrating the results of these three sets of pairs and table 5.1 presents the quantitative results. These results, on the whole, demonstrate what intuition would suggest – that calibrating the model at one location and applying those parameters elsewhere within the same class (green) is preferable to applying those parameters in a similar, but not identical class (yellow) and vastly superior to applying those parameters in an unrelated class (red). The differences between any two clusters (same-class, similar-class, unrelated class) are all significant at the $\alpha = 0.01$ level ($p < .001$ in all cases) as calculated by a two-sample, heteroscedastic t-test (Welch, 1947).

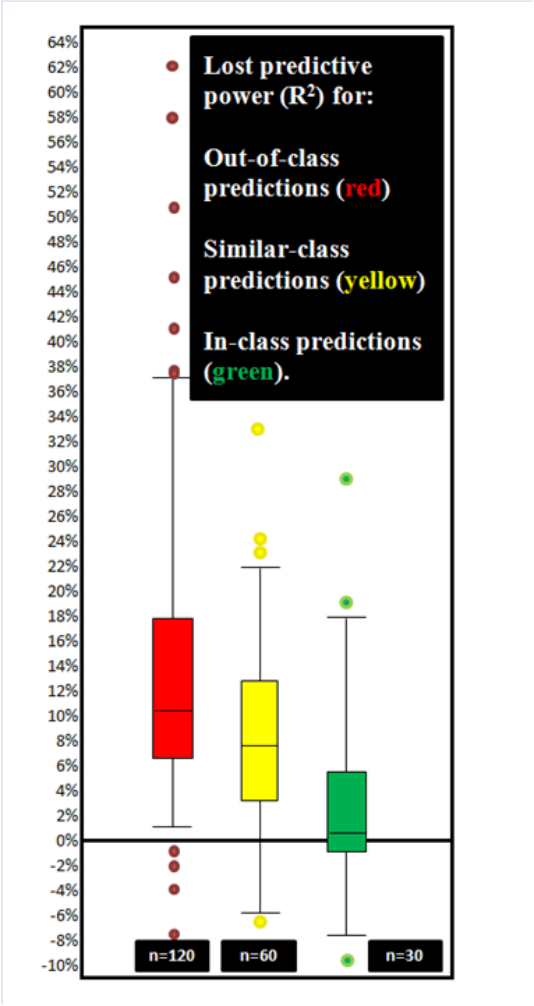


Figure 5.4, Loss of Predictive Power (R^2) (y-axis) Between Baseline Predictions (model calibrated in the same watershed) and Cross-Validation Predictions (model calibrated in other watersheds)

	Unrelated Class	Similar Class	Same Class
Median	-10.5%	-7.3%	-0.8%
Mean	-13.7%	-7.7%	-3.4%
Standard Deviation	1.0%	1.1%	1.4%

Table 5.1, Cross-Validation Results

5.4 Discussion: Generalizing and Improving Models

5.4.1 The Bridge to National Coverage

The results demonstrate that predictions of soil moisture can be made within hydroclimatically similar watersheds without local calibration, a significant step forward. Previous soil moisture models have been calibrated at or near the location at which they will subsequently be applied. Moreover, many of these tools require access to information that is difficult to obtain outside of a research study or, at a minimum, cannot be accessed at numerous locations. By partitioning the nation into similar hydroclimatic classifications using very simple indices (timing of rainfall and runoff), any location in the continental United States, using NEXRAD rainfall and proximally-located USGS streamgauges, can be classified. The results of this analysis demonstrate that a soil moisture model will perform well as long as there is a sensor at which to calibrate the diagnostic soil moisture equation's parameters within the same hydroclimatic class. As only 24 classes describe the entire nation (and only 6 describe a significant majority), it is entirely possible that a couple dozen well-placed soil moisture sensors can enable reasonably accurate soil moisture modeling at any location within the continental United States.

5.4.2 Improvement Over Time: The Value of Machine Learning Enhancement

In addition to generalizing the parameters calibrated in the diagnostic soil moisture equation, the approach presented allows for systematic biases to be removed by searching training data for similar conditions and then predicting the types of mistakes most likely to occur. Figure 5.5, by zooming in upon a 30-day period from figure 5.1, illustrates the value of machine

learning error correction, introducing a diurnal cycle into a model that previously lacked one. The remaining bias is likely explained by a slightly wetter training dataset as compared with the validation data.

By addressing such systematic biases, machine learning enables model performance to improve with each successive growing season as the training dataset expands. For instance, though the fields in Iowa endured flooding during the validation period and subsequently made errors, such errors would eventually populate the training data. The next time such flooding occurs, the model is likely to recognize the occurrence of those same conditions and adjust the diagnostic soil moisture equation's predictions accordingly. In this vein, model performance is likely to improve over time, especially with the models showing strong performance using only a few years of training data.

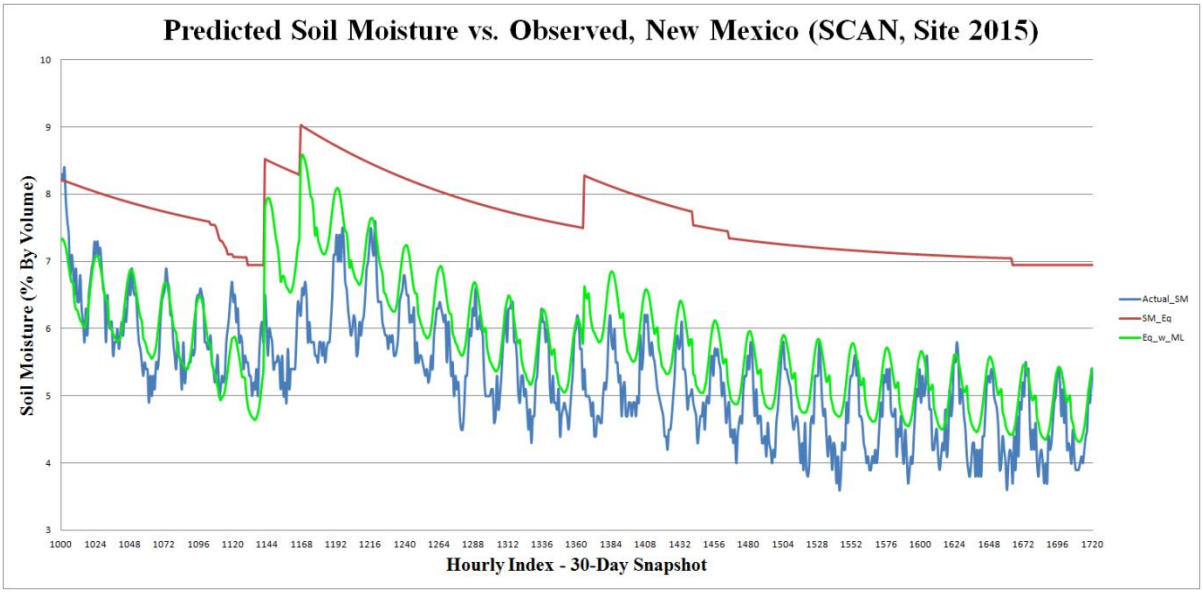


Figure 5.5, Soil Moisture Time Series, SCAN Site 2015, New Mexico (USA), Actual Soil Moisture (Blue Line), Diagnostic Soil Moisture Equation Estimate (Red Line), and Diagnostic Soil Moisture Equation with Machine Learning Error Correction (Green Line)

5.4.3 The Impact of Soils

The fifteen sensors at which the diagnostic soil moisture equation is calibrated and cross-validated are presented in Figure 5.6. Isolating the impacts of soils becomes possible when analyzing groups of sensor locations deemed to be hydroclimatically similar. If cross-application of calibrated parameters yields a notable decline in performance despite similar hydroclimates, then perhaps soil features can help account for this loss in predictive capacity. The soil textural data for each of these fifteen sensors are plotted on a soil texture pyramid diagram in Figure 5.7. These data were obtained from either Pedon Soil Reports available through the SCAN network (which provide precise percentages of clay, silt, and sand), or, where this is unavailable, from soil information in the national soil web database[2].

[2] http://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx

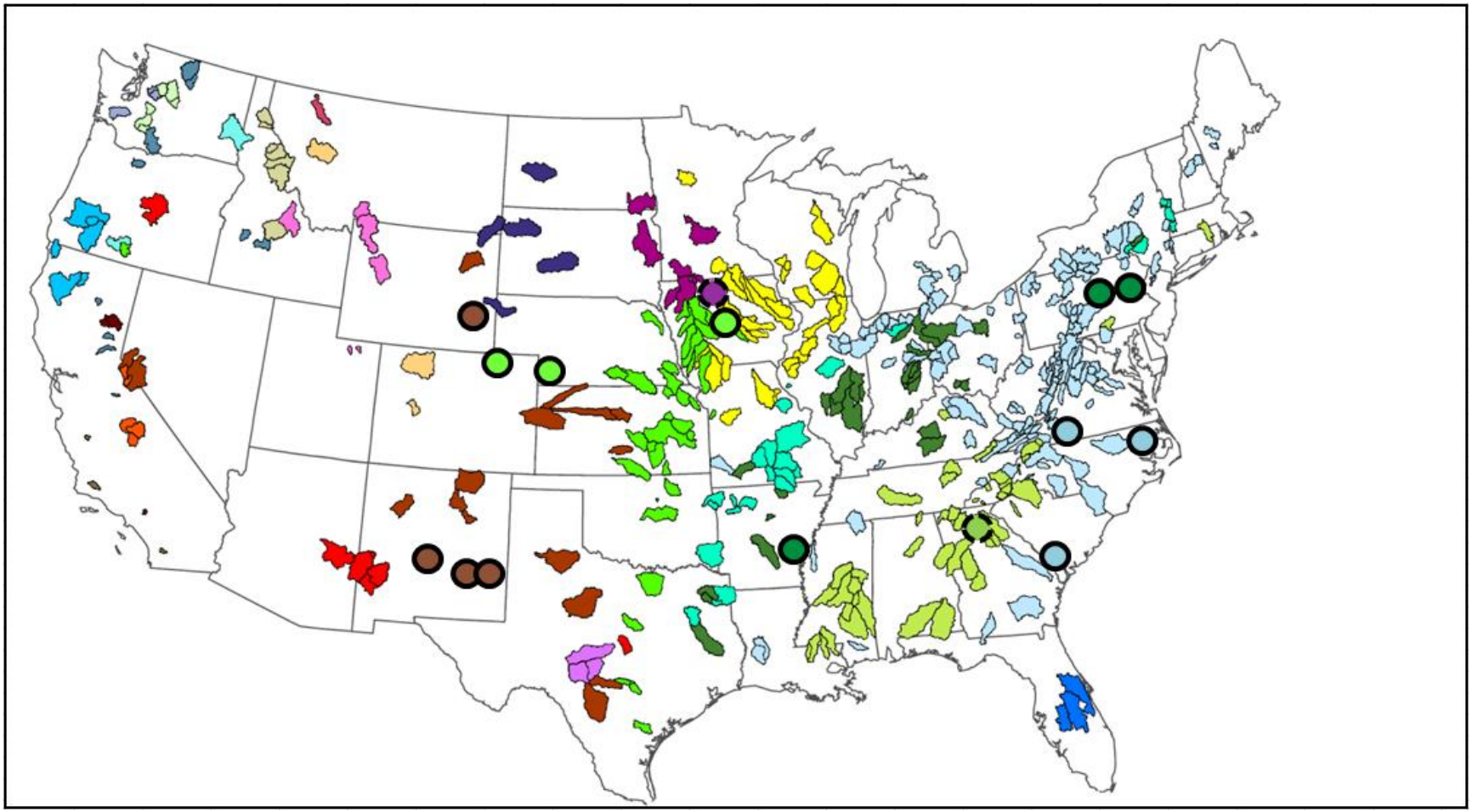


Figure 5.6, 428 MOPEX catchments colored by hydroclimatic class (Coopersmith et al, 2012).
15 SCAN sensors (for which the Diagnostic Soil Moisture Equation is calibrated) are shown as colored circles.
Circle colors correspond to the hydroclimatic class of the point in question.
Circles with dotted borders are unique (no other sensor for calibration is available within that class)

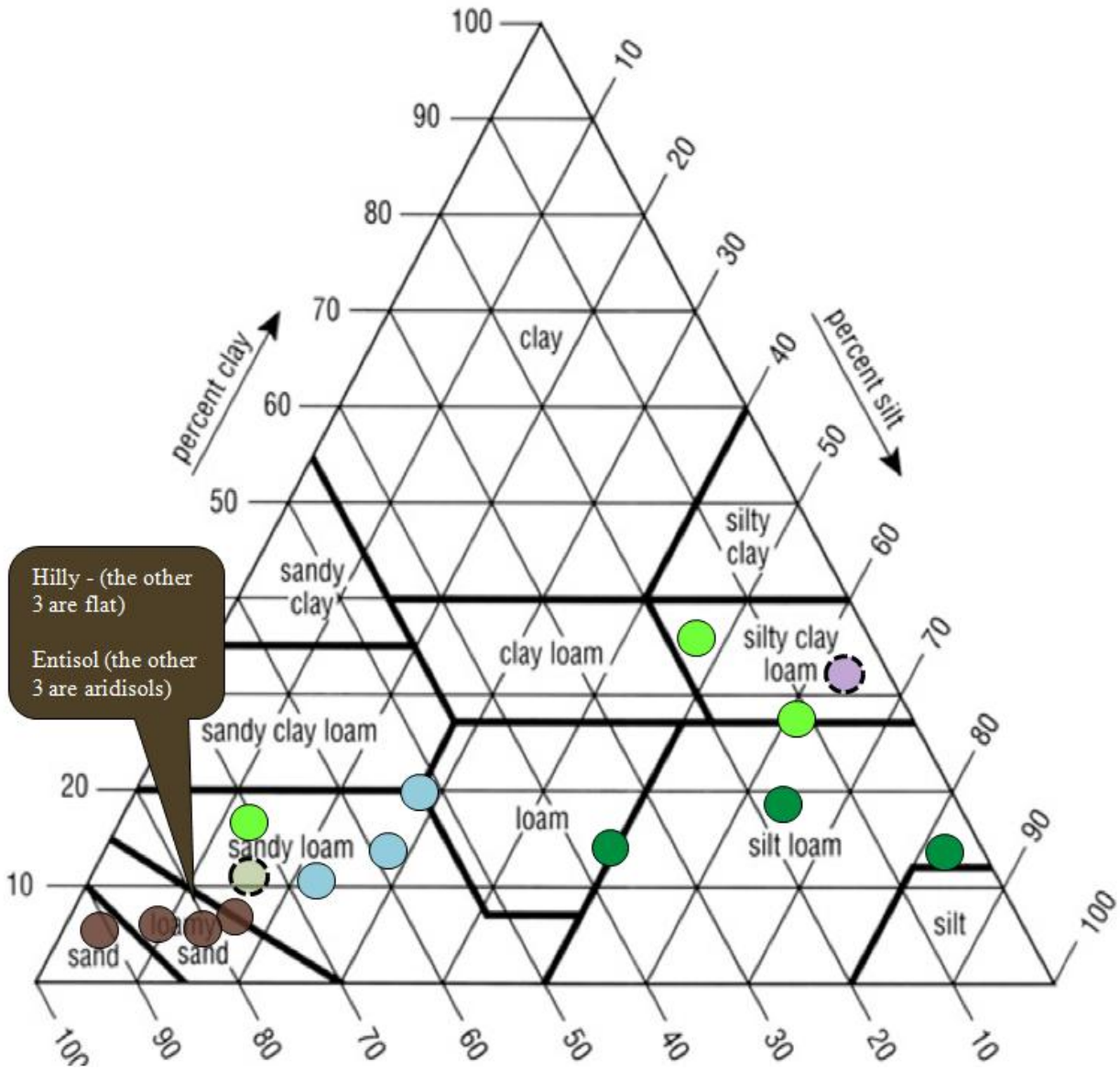


Figure 5.7, The 15 SCAN sensors, color-coded to match their hydroclimatic class.

In figure 5.7, it is noteworthy that of the four hydroclimatic classes for which multiple SCAN sensors are available, three of them (brown, light blue, and dark green) all demonstrate consistency in terms of soil texture as well. Thus, when parameters at one location are shown to be applicable elsewhere within the same class, this is not surprising, as both hydroclimate and soil properties are aligned. However, in the case of the light green dots, two of the three can be

characterized as silty clay loam, but the third falls under the sandy loam designation – a notably different soil. In the case of the brown dots, one location falls in hilly terrain, while the remaining sensors are generally placed on flat ground. These similarities / differences are corroborated by Figure 5.8, which shows that the light blue and dark green classes are all generally ultisols, the light green class consists of two mollisols (the two silty clay loam soils) and one entisol (the sandy loam soil), and the brown class consists of three flat aridisols (in New Mexico) and one hilly entisol (in Wyoming).

Of the thirteen sensors from the four hydroclimatic classes with multiple SCAN sensors (light green, blue, dark green, and brown), 30 (x,y) pairs exist where the model can be calibrated at site x and its parameters applied at site y. Note that (x,y) is not equivalent to (y,x) as the sites for calibration and validation are reversed. Of these 30 pairs, 20 of them are similar in terms of soil and terrain as well. However, 10 of them include a pair of points where the soil types or terrain types are notably misaligned, as noted above. A similar analysis to the one presented in Figure 5.4 and Table 5.1 has been reproduced, comparing the loss in predictive power (R2) for the 20 pairs with similar hydroclimates and soils against the loss for the 10 pairs in which either the soil texture (Figure 5.7) or type (Figure 5.8) do not align. The average loss for the 20 very similar pairs of 1.0% is a much smaller decline than the 8.0% average decline observed for the 10 pairs for which soil/terrain information suggests dissimilarity. These results are significant with a p-value of approximately 0.02. Additionally, the upper-most two green dots in Figure 5.4, where calibrated parameters at one location perform poorly at another of similar hydroclimatic class, fall within these 10 cases.

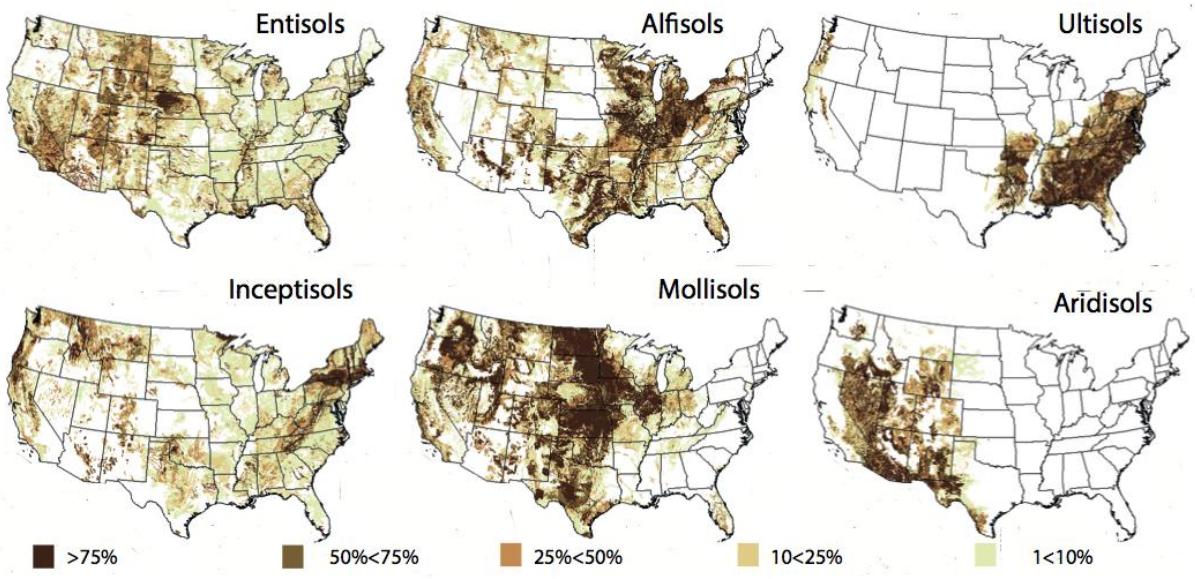


Figure 5.8, Prominent soil taxonomies of the United States (Berghuijs et al, 2013).

These observations show the importance of soil information. While pairs of calibration/validation locations with similar hydroclimates, but dissimilar soils, show a decline in performance as compared with pairs of locations where both are similar, so too do locations with similar soils, but dissimilar hydroclimates. The shaded circles in Figure 5.9 illustrate groups of sensors that are quite similar in terms of soil textures. However, despite their soil similarities, differences in hydroclimates hinder cross-application, showing a decline in performance of 10.9% for all (x,y) pairs within the shaded regions of Figure 5.9 for which x and y are not from the same hydroclimatic class.

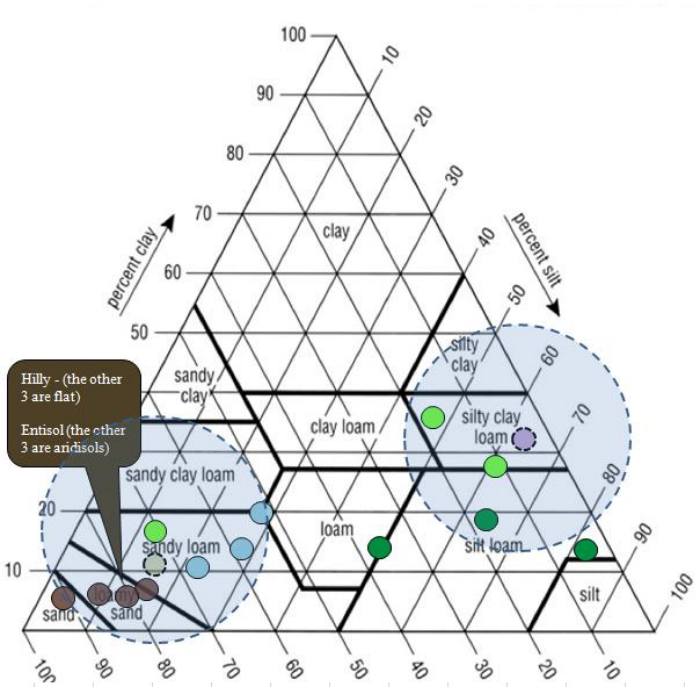


Figure 5.9, The 15 SCAN sensors with similar soil textures shaded.

As summarized in Figure 5.10, these results suggest that in cases where both soil type and hydroclimate align, very little performance is lost when parameters are re-applied (1.0%), moderate declines in performance are observed when one of these two factors are aligned (8.0% if hydroclimates align and soils do not; 10.9% if soils align, but hydroclimates do not), and large declines in performance appear when neither align (20.5%). Clearly both types of attributes are important and should be considered in future modeling work.

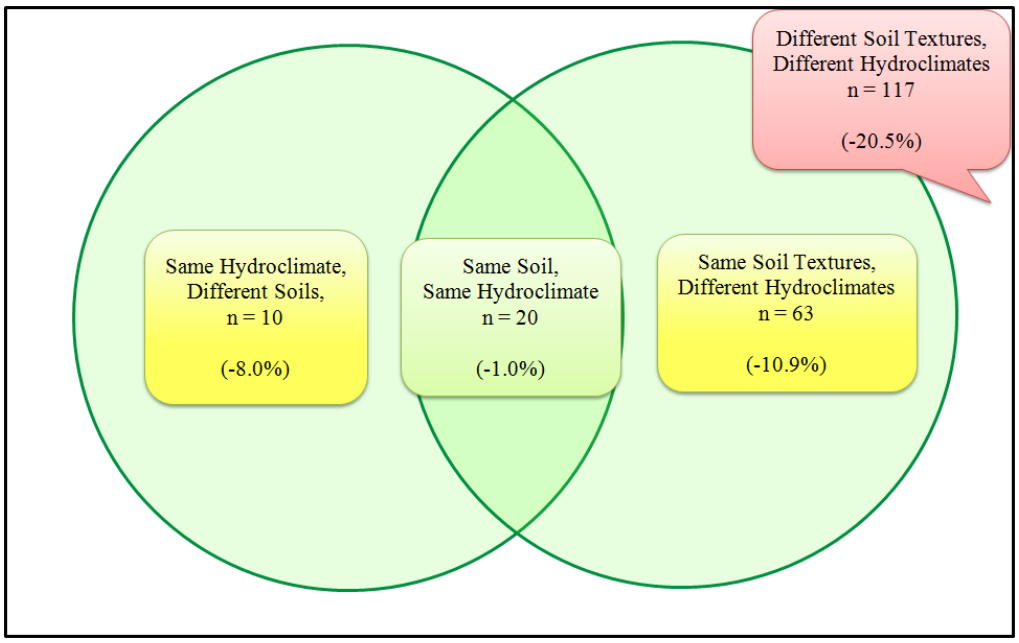


Figure 5.10 Venn-Diagram of Modeling Errors with Similar and Different Soils and Hydroclimates

5.5 Improving Upon the Diagnostic Soil Moisture Equation

5.5.1 Estimates Enhanced By Topographic Classification

Ultimately, the combination of a hydroclimatic classification system and the diagnostic soil moisture equation demonstrates a generalization of calibrations, facilitating predictions at any location where a viable sensor exists within the same hydroclimatic class. One way of improving the accuracy of the approach is to disaggregate the soil moisture estimates as a function of local topography. While SCAN sites used for soil moisture data are generally located on flat surfaces, predictions may be needed at locations located on ridges or in valleys where the soils are likely to be wetter or drier than their surroundings. This requires the notion of regional topological classification. Figure 5.11 shows the topography of a field site located near Ames

IA, modeled at the meter scale using publically-available Light Detection and Ranging (LiDAR) elevation data from the Iowa GeoTree LiDAR Mapping Project[3].

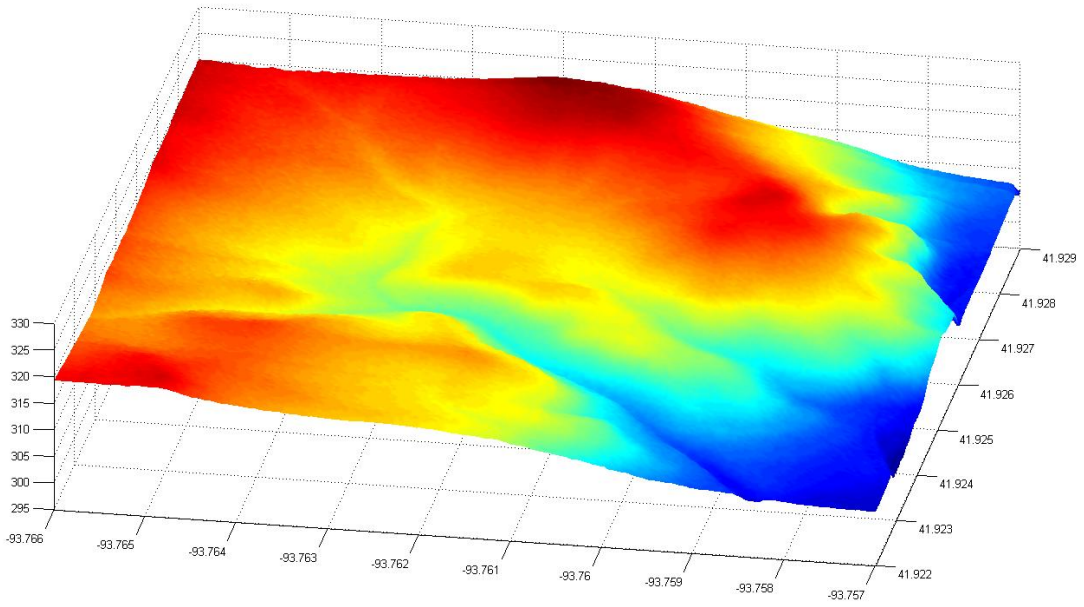


Figure 5.11, Elevations Near Ames, IA. Latitudes (y-axis) and Longitudes (x-axis) Measured in Decimal Form, Elevations (z-axis) Measured in Meters

Simple topographical classifications at three distance scales (1km, 100 m, and 10 m) are presented in Figures 5.12, 5.13, and 5.14, respectively. These three estimates are all calculated by examining the areas surrounding each point at the relevant distance scale and subsequently classifying a location as a ridge, intermediate slope, or valley. Figure 5.15 integrates these three scales, allowing more nuanced shading of the landscape.

[3] http://geotree2.geog.uni.edu/lidar/

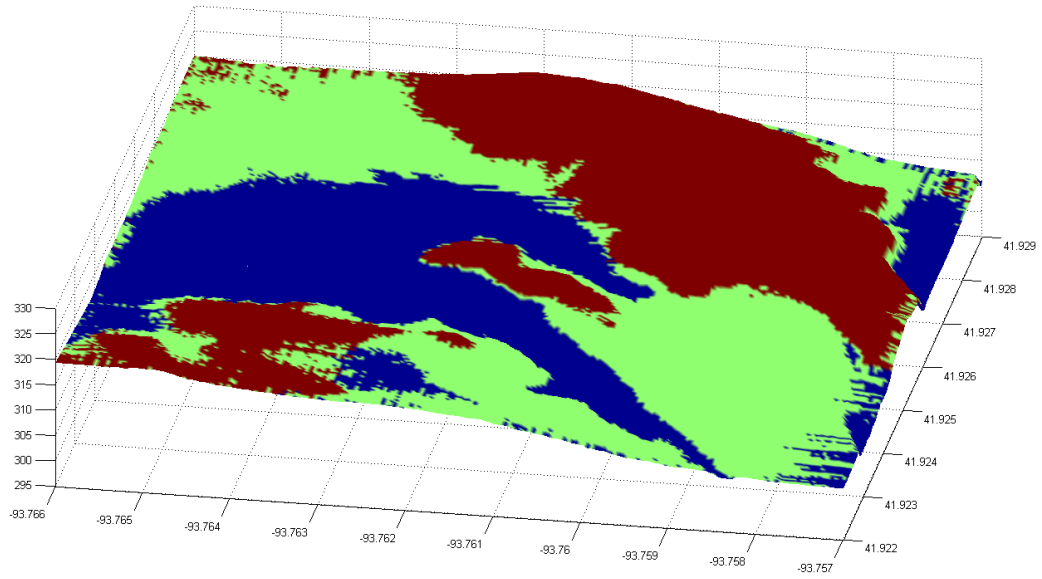


Figure 5.12, Topographic Classification Near Ames, IA (1 km Scale). Ridges Shown in Brown, Valleys in Blue, Intermediate Slopes in Green

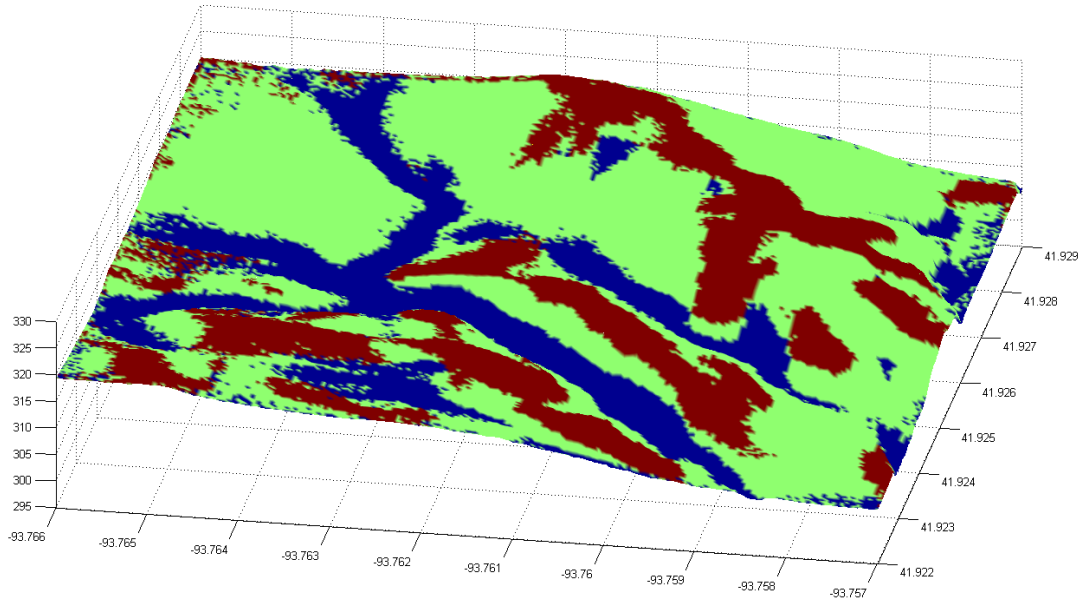


Figure 5.13, Topographic Classification Near Ames, IA (100 m Scale). Ridges Shown in Brown, Valleys in Blue, Intermediate Slopes in Green

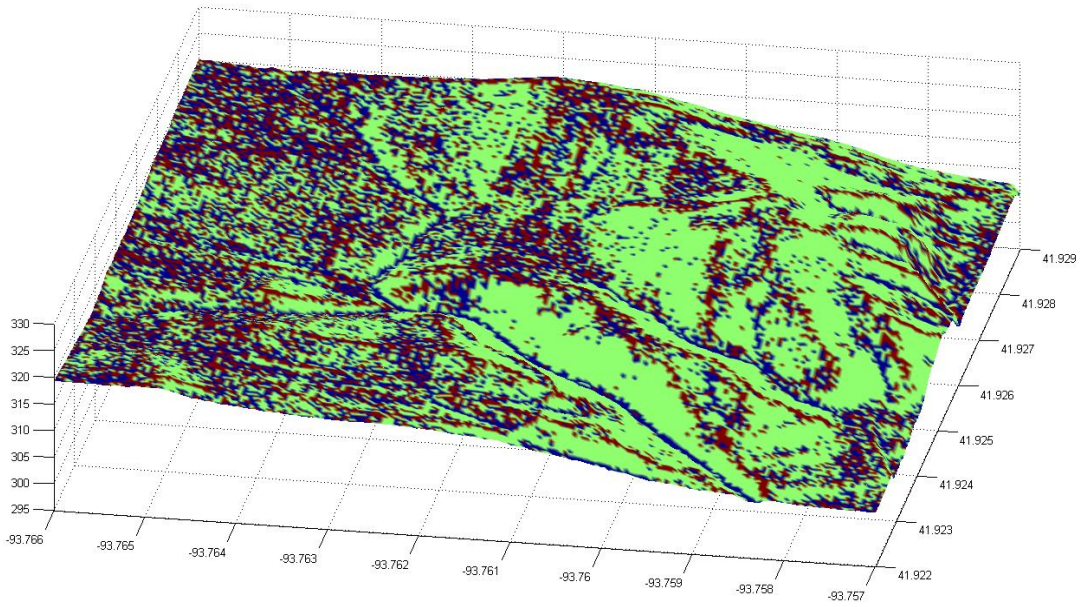


Figure 5.14, Topographic Classification Near Ames, IA (10 m Scale). Ridges Shown in Brown, Valleys in Blue, Intermediate Slopes in Green

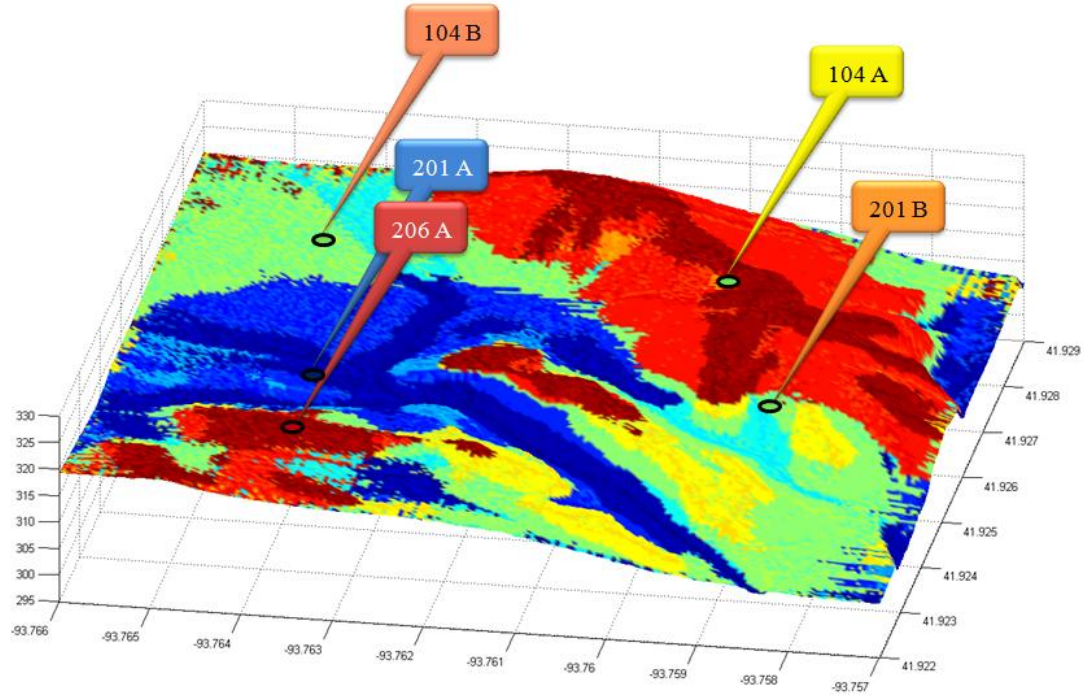


Figure 5.15, Topographic Classification Near Ames, IA (Integration of the Three Scales). Ridges Shown in Brown, Valleys in Blue, Intermediate Slopes in Green

This field site in Iowa contains five soil moisture sensors, as shown in Figure 5.15. The lower-left sensor (206 A, colored red) is located on what is classified as a ridge. The sensor located just north of the ridge (201 A, colored blue) is classified as a valley, while the remaining three sensors (104 A, 104 B, 201 B, colored green) are classified as intermediate slopes.

Soil moisture data at these five sensors in Iowa were recorded from mid-June to early-July of 2012. While rigorous modeling as a function of elevation is not possible with such a limited dataset (only seventeen days exist before the sensor locations were variably tilled), the time series, shown in Figure 5.16, illustrates that elevation plays a substantial role in the soil moisture levels observed. The sensor located in a valley and shown in blue in figure 5.16 gives values that are wetter than the others. The sensor located on a ridge and shown in red in figure 5.16 gives values that are drier than the others. Measurements from the remaining three sensors fall between these upper and lower boundaries.

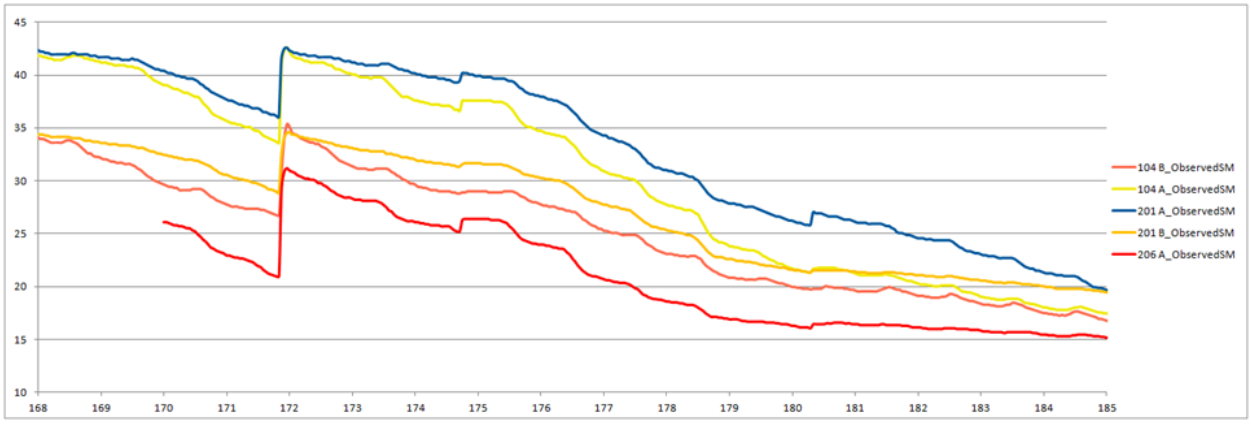


Figure 5.16, Soil Moisture Levels In Percentage By Volume (y-axis) vs. Day of Year (x-axis) Ames, IA (USA)

Future research with more extensive datasets in locations with more complex topological contours should improve soil moisture predictions by enabling the models developed in this work to be adjusted as a function of topographic classification.

5.5.2 An Enhanced Diagnostic Soil Moisture Equation

The diagnostic soil moisture equation could also be improved in future modeling efforts by considering overland flow. Currently, the model assumes that, in the absence of saturation, all rainfall will ultimately infiltrate, as the porosity parameter serves as an upper-bound on the soil moisture levels of the soil. The diagnostic soil moisture equation was designed originally as a daily model, and it is probably rare that on any given day, a significant fraction of precipitation does not infiltrate. However, at the hourly scale it is quite possible that the water from an intense rainfall event will not make its way into the soil at the location of the sensor. In certain Midwestern locations, unusually intense but brief rainfall events may *not* cause increased soil moisture as the diagnostic soil moisture equation would predict. To address this phenomenon, additional parameters can be introduced into the diagnostic soil moisture equation that place an upper bound on the quantity of rainfall that can be considered during any hour (or other interval) of the convolution calculation. While this would require the fitting of additional parameters, it is likely that results would be improved and the concept of variability in soil properties would be described in greater depth than merely porosity, residual soil moisture, and drainage.

5.5.3 NASA's Soil Moisture Active Passive (SMAP) Mission

With NASA satellite data for soil moisture available at the 36 km, 9 km, and 3 km scales throughout the United States and with the SMAP satellite scheduled to launch during the fall of

2014 (O'Neill et al, 2011), the models developed in this work will have ample measurements against which to test and improve their results, and can be used to help check the accuracy of satellite measurements. Future research in LiDAR-driven disaggregation, proposed above, could also be used to improve satellite soil moisture estimates by accounting for smaller-scale topography.

5.6 Conclusions

This work demonstrates the feasibility of providing soil moisture estimates at locations where soil moisture sensors are unavailable for calibration, provided they fall within the same hydroclimatic class (Coopersmith et al, 2012) as a location with a suitable sensor. By calibrating the diagnostic soil moisture equation via a two-part genetic algorithm, improving its performance via a machine learning algorithm for error correction, then validating that algorithm at the same location in subsequent years, a baseline level of predictive performance is established at fifteen locations. Next, these results are cross-validated – deploying parameters calibrated at a given site at sites of similar and different hydroclimatic classes, demonstrating that parameters can be re-applied elsewhere within the same class, but not without.

Leveraging these findings, preliminary analysis is presented, illustrating that soil moisture predictions can be disaggregated based on local topography. This enables more accurate predictions at sites characterized by peaks and valleys that dry faster or slower than the relatively flat locations at which soil moisture algorithms are generally calibrated.

This multi-scale framework, beginning with 428 watersheds throughout the United States, progressing to the field scale with the diagnostic soil moisture equation, and ultimately reaching the meter-scale with finer topographical data from LiDAR, in conjunction with future

satellite sensing missions, demonstrates great promise for producing soil moisture anywhere and anytime.

CHAPTER 6

FUTURE WORK

In this brief chapter, recommendations are made for future work to extend the findings of this dissertation. The discussion begins with a description of how the climate classification work can be expanded to cover the entire nation rather than simply those several hundred catchments for which the system was constructed. Next, an approach for understanding the wetting/drying process at all of these locations will be offered. Finally, these two procedures will be folded together with the results of climate change analysis to form a cohesive modeling approach for hydrologic prediction and decision support.

6.1 Full National Hydroclimatic Classification: Filling in the Gaps

This work has developed a classification system for the 428 catchments contained within the MOPEX database. Though this is a significant step forward, if prediction and decision support are to be offered at any location nationwide, the thousands of catchments not modeled by the classification system must be addressed as well. Recall that the four indices chosen for the classification system can be estimated approximately, even when the exact data used to create the system are incomplete or absent. For example, in many of these catchments, USGS stream gauges provide an estimate of current peak runoff dates (the full 55 years of historical data is no longer necessary given that the tree is already built). Nexrad radar precipitation data offer insight into rainfall timing as well as seasonality estimates. The final variable, the aridity index, can be computed via satellite estimates (Moriyama et al, 2010), from surrounding known catchments, or by integrating remotely sensed Normalized Difference Vegetation Index (NDVI)

data with climate-driven data mining tools (White et al, 2005). While a precise value for Ep/P may not be feasible, the question "is Ep/P > 1?" (Figure 3.11, e.g.) will be answered easily, which is all that the classification ultimately requires.

For any catchment, the four key indices can be calculated or estimated. Where this is not possible, a spatial interpolation can be employed to estimate the class of a given catchment from those surrounding catchments in which a classification is already known.

It is important to acknowledge the limitations inherent in the classification system presented. Although the continental United States represents a diverse and rich array of climate conditions and landscape features, it is natural that it does not contain every conceivable combination of climate and landscapes. It may very well be the case that such climates exist on other continents. It is to be hoped that future efforts will integrate global climate data into an enhanced tree, duplicating this work on a larger, multi-national scale. Secondly, while the classification system classified 428 gauged catchments (including information on runoff timing) into distinct classes, without a further effort to incorporate catchment or landscape features that affect runoff generation, especially runoff timing, application to ungauged catchments is not feasible. This calls for further research that will overcome this limitation.

6.2 Improved Classification Tools: Incorporating Other Features

While the classification tree's four hydroclimatic features have been shown to be valuable in developing soil moisture models that can be re-applied at other locations, it is important to recognize that, were this tree to be reconstructed for the purpose of soil moisture modeling, other parameters could also play an important role. The discussion section from chapter V suggests that to ensure that a soil moisture model can be reapplied elsewhere, it is

important that both the soil properties and hydroclimatic properties of the calibration and validation location are aligned. This could be achieved by adding a feature to the classification tree that incorporates the fractional sand/clay/silt values available through the MOPEX database and elsewhere nationally (SoilWeb, SCAN, e.g.).

Additionally, managerial decisions play a significant role in the wetting and drying processes that govern soil moisture, including crop rotation sequences, tillage schedules, and tile drains. These factors are not available in national datasets and hence were not considered in this research, but provide fertile ground for future analysis and can improve decision support at the local level.

Finally, although this analysis has focused on soil moisture levels near the land surface (depths of 2-4 inches were the focus of this analysis), the depth to the water table is a potentially relevant variable that is not addressed. In areas with shallow water tables, if the water table rises to a depth near the depth of the soil moisture estimate, capillary action may wet the soil even in the absence of rainfall. Water table depths are available at a variety of gauges maintained through the USGS.

Ultimately, a classification tree can incorporate additional layers of data to ensure that two sites that are classified similar in terms of hydroclimatic properties are also similar in terms of soil texture, managerial practice, and drainage properties as well. These features should provide fruitful avenues for future research to improve the classification tree developed in Chapter III.

6.3 Predictions of Soil Moisture from Public Data: LiDAR Disaggregation and Improvements to the Diagnostic Soil Moisture Equation

Another remaining challenge that would enhance the accuracy of soil moisture predictions lies in estimating potential evaporation at any given location. The diagnostic soil moisture equation relies on assumptions of sinusoidal loss functions (the Eta series in chapter V) that may not accurately reflect potential evapotranspiration at any given point in time. While the potential evapotranspiration data available via the Illinois Climate Network were within a mile of the test site at the South Farms in Urbana, IL, such data may not be readily accessible elsewhere where the diagnostic soil moisture equation might prove useful. In these cases, photosynthetically active radiation can be converted to overall solar radiation (Kim and Freyberg, 2011) to then serve as an input into a machine learning algorithm that will return a real-time estimate of the loss function. The algorithm will be flexible enough to allow local refinement via human verifications delivered via the Web, twitter, or mobile application.As discussed in chapter V, to widen the applicability of the diagnostic soil moisture equation, LiDAR elevation data can disaggregate a soil moisture prediction for a given field site, allowing meter-by-meter estimates as a function of topography. While the number of sites with topographically diverse soil moisture sensors is limited, between calibration/validation data from NASA and soil moisture sensor grids made available through the USDA, this line of inquiry seems promising moving forward. Moreover, though LiDAR is currently available in certain states and unavailable in others, it is expected that within the next decade, the entire United States will be mapped at high-resolution by LiDAR data. One might expect the full globe will be available in the years thereafter.

The diagnostic soil moisture equation, though elegant in its minimal data requirements, the relative ease of calibration, and its generalized applicability by hydroclimatic class, can also

be improved by incorporating the notion of infiltration excess as discussed in the previous chapter. It is likely that within certain hydroclimatic classes, the risks of infiltration excess are minimal (catchments with less-intense rainfall and permeable soils). However, within certain classes during seasons in which intense rain events are common, a model adjustment should enhance the quality of results. Similarly, after tillage or other field treatments occur, a field may be rendered temporarily less permeable, a trait that could be suitably addressed by a model that addresses infiltration excess.

Finally, a dashboard of "nowcasts" for soil conditions, viewable as a geospatial browser layer, can be created to allow meaningful visualization of the spatial results for non-experts (e.g., farmers). Additionally, further work is needed to investigate the inclusion of the components of potential evapotranspiration (which came from the Illinois State Water Survey) from national sources, in order to expand wetting/drying assessments in Chapter II beyond the state of Illinois. It is worth noting that while precipitation data are becoming increasingly ubiquitous globally, potential evaporation estimates could be problematic outside the United States. However, remotely sensed solar radiation from satellites and other sources may ultimately resolve this issue.

6.4 Hydroclimatic Change Projection and Impacts: The Future National Climate

Once the entire nation is classified and soil moisture can be estimated at any desired location, one final piece to the puzzle will remain. The climate classes, as they currently stand, are unlikely to endure in subsequent years and decades. Extrapolating somewhat from the regime shifts noted over the past 20-30 years, it is possible to infer the ultimate classification of

catchments in future contexts. This will help develop the types of machine learning algorithms required for decision support in the future.

Such algorithms can provide concrete guidance into various land-use decisions under the purview of local managers. As the climate changes, the need for irrigation may increase in one region and become unnecessary in another. With lengthened or shortened growing seasons, the timing or selection of crops may shift. By tracking how climate classifications are shifting, the agricultural community can be prepared for such changes and thereby minimize adverse impacts.

Though ultimately, decisions are made by human intuition rather than the prescriptions of computation-driven models, this research demonstrates that the combination of local data for machine learning models of the wetting/drying process, a national-climate classification system, and an understanding of regional climate change can aid decision support over a broad spatial and temporal space.

REFERENCES

Anthes, R.A., Corell, R.W., Holland, G., Hurrell, J.W., MacCracken, M.C., Trenberth, K.E., "Hurricanes and global warming – Potential linkages and consequences," Bull. Am. Meteorol. Soc. 87, 623–628. 2006. doi:10.1175/BAMS-87-5-617

Aertsen, W., Kint, V., Van Orshoven, J.,and Muys, B. "Evaluation of modeling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA)." Environmental Modelling and Software, Volume 26, Issue 7, July 2011, Pages 929-937.

Barricelli, N.A. "Numerical testing of evolution theories. Part II. Preliminary tests of performance, symbiogenesis and terrestrial life". *Acta Biotheoretica* (16): 99–126. 1963.

Beckage, B., Osborne, B., Gavin, D.G., Pucko C., Siccama T., Perkins, T. "A rapid upward shift of a forest ecotone during 40 years of warming in the Green Mountains of Vermont." Proc. Natl. Acad. Sci. 105:4197-4202 (2008).

Beniston, M., (2003), Climatic change in mountain regions: A review of possible impacts. Climatic Change, 59 (1-2), pp. 5-31, doi: 10.1023/A:1024458411589.

Berghuijs, W.R., Sivapalan, M., Savenije, H.H.G., and Woods, R.A. "The seasonal water balance as a window to explore catchment similarity at various time-scales." In Progress, 2013.

Blanchard, B. J., Mcfarland, M. J., Schmugge, M. J., and Rhoades, E. "Estimation of soil-moisture with API algorithms and microwave emission," Water Resour. Bull., 17, 767– 774, 1981.

Bloschl, G. and Sivapalan, M., "Scale issues in hydrological modeling: a review." Hydrological Processes. 9 (3-4), pp. 251-290.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA. 1984.

Breiman, L., et al., *Classification and Regression Trees*, Chapman & Hall, Boca Raton, 1993.

Brodley, C.E., Freidl, M.A. "Decision tree classification of land cover from remotely sensed data." Remote Sensing of Environment, Volume 61, Issue 3, September 1997, Pages 399-409.

Budyko, M.I. Climate and Life, Academic, New York, 1974.

Burn, D. H. (1997), Catchment similarity for regional flood frequency analysis using seasonality measures, Journal of Hydrology, 202(1-4), 212-230, doi: 10.1016/s0022-1694(97)00068-1.

Burn, D. H., and N. K. Goel (2000), The formation of groups for regional flood frequency analysis, Hydrological Sciences Journal, 45(1), 97-112, doi: 10.1080/02626660009492308.

Cai, X., Wang, D., Zhu, T., and Ringler, C. “Assessing the regional variability of GCM Simulations.” Geophys. Res. Lett. 36 L02706, 2009.

Cai, X., Zhang X., and Wang, D. “Land Availability for Biofuel Production.” Environ. Sci. Technol., 2011, 45 (1), pp 334–339, DOI: 10.1021/es103338e

Capehart, W. J., and Carlson, T.N. “Estimating near-surface soil moisture availability using a meteorologically driven soil water profile model,” J. Hydrol., 160, 1– 20, 1994.

Carrillo, G., Troch, P. A., Sivapalan, M., Wagener, T., Harman, C., and Sawicz, K. (2011). Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient, *Hydrol. Earth Syst. Sci.*, 15, 3411-3430, doi:10.5194/hess-15-3411-2011

Cayan, D.R., Kammerdiener, S.A., Dettinger, M.D., Caprio, J.M., and Peterson, D.H. (2001), Changes in the onset of spring in the western United States. 82(3), 2001, pp. 399-415.

Cheng, L., Yaeger, M., Viglione, A., Coopersmith, E., Ye, S., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 1: Insights from statistical analyses, Hydrol. Earth Syst. Sci. Discuss., 9, 7001–7034, doi:10.5194/hessd-9-7001-2012, 2012.

Chico-Santamarta, L., Richards, T., and Godwin, R.J. "A laboratory study into the mobility of travelling irrigators in air dry, field capacity and saturated sandy soils." American Society of Agricultural and Biological Engineers Annual International Meeting 2009. Volume 4, 2009, Pages 2629-2646.

Choudhury, B. J. and Blanchard, B.J., "Simulating soil water recession coefficients for agricultural watersheds," Water Resour. Bull., 19, 241–247, 1983.

Coopersmith, E.J., Minsker B., and Montagna P. "Understanding and forecasting hypoxia using machine learning algorithms." Journal of Hydroinformatics, Volume 13, Issue 1, 2011. Pages 64-80.

Coopersmith, E., Yaeger, M., Ye, S., Cheng, L., and Sivapalan, M. "Exploring the physical controls of regional patters of flow duration curves – Part 3: A catchment classification system based on seasonality and runoff regime, Hydrol. Earth Syst. Sci. Discuss., 9, 7085-7129, doi: 10.5194/hessd-9-7085-2012, 2012.

Dooge, J. I. C., "Looking for hydrologic laws," Water Resources Research, Vol. 22, No. 9S, PP. 46S-58S, 1986.

Drummond, M.A., Auch, R.F., Karstensen, K.A., Sayler, K. L., Taylor, J.L., Loveland, T.R. "Land change variability and human-environment dynamics in the United States Great Plains," Land Use Policy 29 (2012) 710-723.

Duan, Q., Sorooshian, S., and Gupta, V.K., "Effective and efficient global optimization for conceptual rainfall-runoff models." Water Res. R., 28, 1015-1031, 1992.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H., Gusev, Y., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, J Hydrol, 320, 3-17, 2006.

Entekhabi, D. and Rodriguez-Iturbe, I., "Analytical framework for the characterization of the space-time variability of soil moisture," Adv. Water Resour., 17, 35–45, 1994.

Evans, A.M., and Perschel, R. "A review of forestry mitigation and adaptation strategies in the Northeast U.S." Climatic Change 96:167-183 (2009).

Farago, T., "Soil moisture content: Statistical estimation of its probability distribution," J. Clim. Appl. Meteorol., 24(4), 371– 376, 1985.

Fix, E., Hodges, J.L. Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

Foster, D.R., Clayden, S., Orwig, D.A., Hall, B., Barry, S. “Oak, chestnut and fire: climate and cultural controls of long-term forest dynamics in New England, USA.” Jour. Biogeogr. 29: 1359-1379. (2002).

Freund, Y. and Schapire, R.E. “A decision-theoretic generalization of on-line learning and an application to boosting.” Journal of Computer and System Sciences, no. 55. 1997

Gamache, R.W., Kianirad, E., and Alshawabkeh, A.N. “An automatic portable near surface soil characterization system.” Geotechnical Special Publication, Issue 192, 2009, Pages 89-94.

Goldenberg, S.B, Landsea, C.W., Mestas-Nunez, A.M., Gray, W.M. “The recent increase in Atlantic hurricane activity: Causes and implications,” Science 20 July 2001: Vol. 293 no. 5529 pp. 474-479 DOI: 10.1126/science.1060040.

Gong, J. and Caldas, C.H. “Data processing for real-time construction site spatial modeling.” Automation in Construction, Volume 17, Issue 5, July 2008, Pages 526-535.

Haines, A. T., Finlayson, B. L., and McMahon, T. A., (1988) A global classification of river Regimes, Applied Geography, 8, 255-272.

Hargreaves, M.V.M., 1993. “Dry Farming in the Northern Great Plains Year of Readjustment, “1920-1990. University Press of Kansas, Lawrence, KS. 396pp.

Harmon, R.W. (1963). “Computation of direct runoff amounts from storm rainfall.” International Association of Scientific Hydrology. Publication 63, Wallingford, Oxon., U.K.

Haywood, A., Stone C., “Mapping eucalypt forest susceptible to dieback associated with bell miners (Manorina melanophys) using laser scanning, SPOT 5, and ancillary topographical data.” Ecological Modelling 222 (5), pp. 1174-1184, 2011.

Jackson, T.J., Bindlish, R., Cosh, M.H., Zhao, T., Starks, P.J., Bosch, D.D., Seyfried, M., Moran, M.S., Goodrich, D.C., Kerr, Y.H., Leroux, D., “Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the U.S.” IEEE Transactions on Geoscience and Remote Sensing, Volume 50, Issue 5, Part 1, 1530-1543. May 2012.

Jensen, M. E, Burman, R. D., and Allen, R. G. (1990) “Evapotranspiration and irrigation water requirements.” Manuals and Reports of Engineering Practice, No.70. New York.

Jog, G.M., Brilakis, I.K., and Angelides, D.C. “Testing in harsh conditions: Tracking resources on construction sites with machine vision.” Automation in Construction Volume 20, Issue 4, July 2011, Pages 328-337.

Jog, G.M., Brilakis, I.K., and Angelides, D.C. “Testing in harsh conditions: Tracking resources on construction sites with machine vision.” Automation in Construction Volume 20, Issue 4, July 2011, Pages 328-337.

Jones, H. G. (2004). “Irrigation scheduling: advantages and pitfalls of plant-based methods.” Journal of Experimental Botany, 55, 2427-2436.

Kim, J.Y., and Freyberg, D.L. “Using Photosynthetically Active Radiation (PAR) Observation to Estimate Potential Evaporation with Combination Equations.” Poster Session H33A. AGU Fall Meeting, 2011.

Kim, S., Shin, H., Joo, K., Heo, J.H. “Development of plotting position for the general extreme value distribution.” Journal of Hydrology, Volume 475, 259-269. December 2012.

Koppen, W. Das geographisca System der Klimate. Handbuch der Klimatologie, edited by: Koppen, W. and Geiger, G. 1. C. Gebr, Borntraeger, 1-44, 1936.

Kropp, A. “Survey of residential foundation design practice on expansive soils in the San Francisco bay area.” J. Perf. Constr. Fac. 25, 24 (2011); doi:10.1061/(ASCE)CF.1943-5509.0000074 (7 pages)

Kumar, P., Alameda, J., Bajcsy, P., Folk, M., and Markus, M. Hydroinformatics. CRC Press, Taylor & Francis Group, New York, USA. 2006.

Lailiberte, A.S., Fredrickson, E.L., and Rango, A. “Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands.” Photogrammetric Engineering and Remote Sensing, Volume 73, Issue 2, February 2007, Pages 197-207.

Lamande, M. and Schjonning, P. "The ability of agricultural tyres to distribute the wheel load at the soil-tyre interface." Journal of Terramechanics 45 (4), pp. 109-120. 2008.

Lebert, M., Brunotte, J., Sommer, C., and Boken, H. "Protecting soil structure against compaction: Proposed solutions to safeguard agricultural soils." Journal of Plant Nutrition and Soil Science 169 (5), pp. 633-641. 2006.

Lee, J.H. and Wang, W. "Characterization of snow cover using ground penetrating radar for vehicle trafficability - Experiments and modeling." Journal of Terramechanics, Volume 46, Issue 4, August 2009, Pages 189-202.

Li, H., Sun, J., Wu, J. "Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods." Expert Systems with Applications 37 (2010) 5895-5904.

Li, Y. Jiao, Y., and Reid, K. "Assessment of landed and non-landed by-catch of walleye, yellow perch, and white perch from the commercial gillnet fisheries of Lake Erie, 1994-2007." Journal of Great Lakes Research, In-Press.

Littell, J.S., Elsner, M.M., Binder, L.C.W., and Snover, A.K., "The Washington Climate Change Impacts Assessment: Evaluating Washington's Future in a Changing Climate*."* Climate Impacts Group, University of Washington, Seattle, Washington. 2009.

Lobell, D.B., Hammer, G.L., McLean, G., Messina, C., Roberts, M.J., Schlenker, W. "The critical role of extreme heat for maize production in the United States" *Nature Climate Change*, Letter (2013). doi:10.1038/nclimate1832

McDonnell, J.J., and Woods, R. "On the need for catchment classification." Journal of Hydrology. 299 (1-2), pp. 2-3. 2004

McRoberts, R., Tomppo, E., Finley, A., and Heikkinen, J. "Estimating areal means and variances of forest attributes using the k-nearest technique and satellite imagery." Remote Sensing Environ. 111 (4), 466-480. 2007.

Meliker, J.R. and G.A. AvRuskin, M.J. Slotnick, P. Goovaerts, D. Schottenfeld, G.M. Jacquez, and J.O. Nriagu. "Validity of spatial models of arsenic concentrations in private well water." Environmental Research; Jan2008, Vol. 106 Issue 1, p42-50.

Milly, P. C. D.: Climate, soil water storage, and the average water balance, *Water Resour. Res.*, 30, 2143-2156. 1994.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., and Stouffer, R.J. "Stationarity is Dead: Whither Water Management?" *Science* 1, February 2008, Vol. 319 no. 5863, pp. 573-574.

Monteith, J. L. (1965). “Evaporation and environment.” In Proceedings of the 19th symposium of the Society for Experimental Biology. 205-233, Cambridge Universality Press, New York.

Moriyama, M., Tanigawa, S., and Kimura, R. “Satellite based aridity mapping.” International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, Volume 38, Part 8, Kyoto, Japan, 2010.

Mosaddeghi, M.R., Hajabbasi, M.A., Hemmat, A., and Afyuni, M. “Soil compactibility as affected by soil moisture content and farmyard manure in central Iran.” Soil and Tillage Research 55 (1-2), pp. 87-97. 2000.

Mosley, M. P. (1981), Delimitation of New Zealand hydrologic regions, Journal of Hydrology, 49(1-2), 173-192, doi: 10.1016/0022-1694(81)90211-0.

Munson, S.M., Webb, R.H., Belnap, J., Hubbard, A.H., Swann, D.E., and Rutman, S. “Forecasting climate change impacts to plant community composition in the Sonoran Desert region.” Global Change Biology, Volume 18, Issue 3, pp. 1083-1095, March 2012.

Nemes, A., Roberts, R., Rawls, W., Pachepsky, Y., and Van Genuchten, M. “Software to estimate – 33-1500 kPa soil water retention using the non-parametric k-nearest-neighbor technique” Environ. Modell. Softw. 23(2) 254-255. 2008.

Nolin, A.W. (2012), Perspectives on climate change, mountain hydrology, and water resources in the Oregon Cascades, USA, Mountain Research and Development. Volume 32, Issue Suppl. 1, March 2012, pp. S35-s46.

Ogunkoya, O. O. (1988), Towards a delimitation of southwestern Nigeria into hydrological regions, Journal of Hydrology, 99(1-2), 165-177, doi: 10.1016/0022-1694(88)90085-6.

Olden, J., Kennard, M., and Pusey, B.: A framework for hydrologic classification with a review of methodologies and applications in ecohydrology, Ecohydrol. 2011.

O'Neill, P., Entekhabi, D., Njoku, E, and Kellogg, K. "The NASA Soil Moisture Active Passive (SMAP) Mission: Overview". *NASA*. Goddard Space Flight Center, Jet Propulsion Laboratory. Retrieved 14 May 2013. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110015242_2011016052.pdf

Pan, F., "Estimating daily surface soil moisture using a daily diagnostic soil moisture equation. Journal of Irrigation and Drainage Engineering, 138(7), 625-631, 2012.

Pan, F., and Peters-Lidard, C. D. (2008). "On the relationship between the mean and variance of soil moisture fields." Journal of the American Water Resources Association. 44(1), 235-242.

Pan, F., Peters-Lidard, C. D., and Sale, M. J. (2003). "An analytical method for predicting surface soil moisture from rainfall observations." Water Resources Research, 39(11), Art.

Parshall, T., Foster, D.R., Faison, E., Macdonald, D., Hansen, B.C.S., "Long-term history of vegetation and fire in pitch pine-oak forests on Cape Cod, Massachusetts." Ecology 84:736-748. (2003).

Pena-Barragan, J.M., Ngugi, M.K., Plant R.E., and Six, J. "Object-based crop identification using multiple vegetation indices, textural features and crop phenology." Remote Sensing of Environment, Volume 115, Issue 6, 15 June 2011, Pages 1301-1316.

Peel, C.M., Finlayson, B.L., and McMahon, T.A. Updated world map of the Koppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.,* 11, 1633-1644, 2007.

Pinto, A., Nunes, I.L., and Ribeiro, R.A. "Occupational risk assessment in construction industry – overview and reflection." Safety Science, Volume 49, Issue 5, June 2011, Pages 616-624.

Priestley, C. H. B., and Taylor, R. J. (1972). "On the assessment of surface heat flux and evaporation using large-scale Parameters." Monthly Weather Review, 100: 81-92.

Pytka, J. "Determining and analyzing the stress state under wheeled-vehicle loads." Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering Volume 223, Issue 2, 1 February 2009, Pages 233-253.

Quinlan, J. R. 1986. Induction of Decision Trees. *Mach. Learn.* 1(1),, 81-106, 1986.

Raper, R.L. "Agricultural traffic impacts on soil." Journal of Terramechanics 42 (3-4), pp. 259-280. 2005.

Rechenberg, I. Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Frommann-Holzboog Verlag, Stuttgart, 1973.

Russell S., and Norvig, R. Artificial Intelligence: A Modern Approach. Prentice Hall, Indianapolis, 2010.

Sahu, R.K. and Raheman, H. "A decision support system on matching and field performance prediction of tractor-implement system." Computers and Electronics in Agriculture, 60 (1), pp. 76-86. 2008.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G. (2011). Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sci.*, 15, 2895-2911, doi:10.5194/hess-15-2895-2011

Saxton, K. E., and Lenz, A.T, "Antecedent retention indexes predict soil moisture," J. Hydraul. Div. Proc. Am. Soc. Civ. Eng., 93, 223– 241, 1967.

Schmugge, T., Gloersen, P., Wilheit, T., and Geiger, F. “Remote sensing of soil moisture with microwave radiometers.” *Jour. of Geophys. Res.* 79 (2), 317-323, Jan. 1974.

Schwefel, H.P. “Evolutionsstrategie und numerische Optimierung.” Dissertation, TU Berlin, Germany, 1975.

Seibert, J., and Vis. M.J.P., “Teaching hydrological modeling with a user-friendly catchment-runoff-model software package.” *Hydrol. Earth Syst. Sci.* 16 (9), 3315-3325, 2012.

Sharifat, K., and Kushwaha, R.L., “Sinkage simulation model for vehicles on soft soil.” 2000 ASAE Annual Intenational Meeting, Technical Papers: Engineering Solutions for a New Century 1, pp. 2549-2553

Simunek J., Sejna, M., van Genuchten, M. “The HYDRUS-1D software package for simulating water flow and solute transport in two-dimensional variably saturated media. Version 2.0, IGWMC –TPS –70, International Ground Water Modeling Center, Colorado School of Mines, Golden, CO, 1998.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Menidiondo, E.M., O’Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., and Zehe, E. IAHS decade on predictions in Ungauged Basins. (PUB), 2003-2012. Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, 48, 857-880, 2003.

Sivapalan, M., Yaeger, M.A., Harman, C.J., Xu, X. and Troch, P.A."Functional model of water balance variability at the catchment scale: 1. Evidence of hydrologic similarity and space-time symmetry," *Water Resour. Res., 47*, W02522, 1992 doi:10.1029/2010WR009568

Shepard, D. "A two-dimensional interpolation function for irregularly-spaced data." Proceedings of the 1968 23rd ACM national conference, pp 517-524, 1968. 10.1145/800186.810616.

Shoop, S., Kestler, M., Stark, J., Ryerson, C., Affleck, R. "Rapid stabilization of thawing soils: Field experience and application." Journal of Terramechanics 39 (4), pp. 181-194. 2002.

Sliva, R.B., Lancas, K.P., Miranda, E.E.V., Silva, F.A.M., and Baio, F.H.R. "Estimation and evaluation of dynamic properties as indicators of changes on soil structure in sugarcane fields of Sao Paulo State – Brazil." Soil and Tillage Research 103 (2), pp. 265-270. 2009.

Stam, J.M., Dixon, B.L. "Farmer Bankruptcies and Farm Exits in the United States, 1899-2002." Economic Research Service, U.S. Department of Agriculture, Agriculture Bulletin Number 788, Washington D.C. 2004.

Stavrakoudis, D.G., Theocharis, J.B., and Zalidis, G.C. "A boosted genetic fuzzy classifier for land cover classification of remote sensing imagery." ISPRS Journal of Photogrammetry and Remote Sensing, In-Press.

Tang, G., Beckage, B., and Smith, B. “The potential transient dynamics of forests in New England under historical and projected future climate change.” DOI 10.1007/s10584-012-0404-x.

Thornthwaite, C.W. (1948). “An approach toward a rational classification of climate.” Geographical Review, 38: 55-94.

Todhunter, P.E., Rundquist, B.C., 2004. “Terminal lake flooding and wetland expansion in Nelson County, North Dakota,” Physical Geography 25, 68-85.

Trenberth, K.E., “Changes in precipitation with climate change.” Climate Research, Vol. 47, 123-138, 2011. doi: 10.3354/cr00953

Tullberg, J.N, Yule, D.F., McGarry, D. “Controlled traffic farming: From research to adoption in Australia.” Soil and Tillage Research 97 (2), pp. 272-281. 2007.

Urban, D., Roberts, M.J., Schlenker, W., Lobell, D.B. “Projected temperature changes indicate significant increase in interannual variability of U.S. maize yields.” Climatic Change, 2012. DOI 10.1007/s10584-012-0428-2

Van Hooff, T., Blocken B., and Van Harten, M. “3D CFD simulations of wind flow and wind-driven rain shelter in sports stadia: Influence of stadium geometry.” Building and Environment, Volume 46, Issue 1, January 2011, Pages 22-37.

Wagener, T., Sivapalan, M., Troch, P. A., and Woods, R. A.: Catchment Classification and Hydrologic Similarity, *Geog. Comp.,* 1/4, 901-931. 2007.

Walsh, R.P.D. and Lawler, D.M. Rainfall seasonality: description, spatial patterns and change through time (British Isles, Africa). *Weather*, 36(7), 201-208, 1981.

Wang, D. and Hejazi, M. "Quantifying the relative contribution of the climate and direct human impacts on mean annual streamflow in the contiguous United States." Water Resources Research, Vol. 47, doi:10.1029/2010WR010283, 2011.

Weare, B.C. and Blossier, B., "Snow water changes with elevation over Western United States in IPCC AR4 models" DOI 10.1007/s10584-011-0258-7

Welch, B. L. (1947). "The generalization of "Student's" problem when several different population variances are involved". Biometrika 34 (1–2): 28–35. doi:10.1093/biomet/34.1-2.28

Wetzel, P. J., and Chang, J. T., "Evapotranspiration from nonuniform surfaces— A 1st approach for short-term numerical weather prediction," Mon. Weather Rev., 116, 600–621, 1988.

White, A.B., Kumar, P., and Tcheng, D. "A data mining approach for understanding topographic control on climate-induced inter-annual vegetation variability over the United States." Remote Sensing of Environment, Vol. 98, Issue 1, 1-20, 2005.

Woodward, F.I. and Beerling, D.J., "The dynamics of vegetation change: health warnings for equilibrium 'dodo' models." Global Ecol. Biogeogr. 6:413-418. (1997).

Yadav, M.T., Wagener, and Gupta, H. V. Regionalization of constraints on expected watershed response. Adv. Water Resour., 30, 1756-1774, 2007.

Yaeger, M., Coopersmith, E., Ye, S., Cheng, L., Viglione, A., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 4: A synthesis of empirical analysis, process modeling and catchment classification, Hydrol. Earth Syst. Sci. Discuss., 9, 7131–7180, doi:10.5194/hessd-9-7131-2012, 2012.

Ye, S., Yaeger, M. A., Coopersmith, E., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 2: Role of seasonality and associated process controls, Hydrol. Earth Syst. Sci. Discuss., 9, 7035–7084, doi:10.5194/hessd-9-7035-2012, 2012.

Yokoo, Y, and Sivapalan, M. (2011). Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis. HESS.

Zhang L., Dawes, W.R., and Walker, G.R. "Response of mean annual evapotranspiration to vegetation changes at catchment scale." Water Resources Research 37 701-708. 2001.

Zhang, X., and Cai, X. “Climate change impacts on global agricultural land availability.” Environ. Res. Lett. 6 (2011).

APPENDIX A

JUSTIFICATION OF THE FOUR INDICES FROM CHAPTER 3

The primary purpose of this supplement is to provide further, more detailed justification of the four indices chosen for the classification tree presented in the full manuscript.

Figure A.1 demonstrates the superior descriptive capacity of the tree with all four indices included as compared with the tree that is created from the exclusion from each of the four variables. Four separate trees are created (each uses a different combination of three of the four indices), and the variance at each level of the tree is measured. When these results are compared with the decrease in variance by employing all four indices, it is clear that none of these four indices can be removed without a substantial loss of descriptive power.

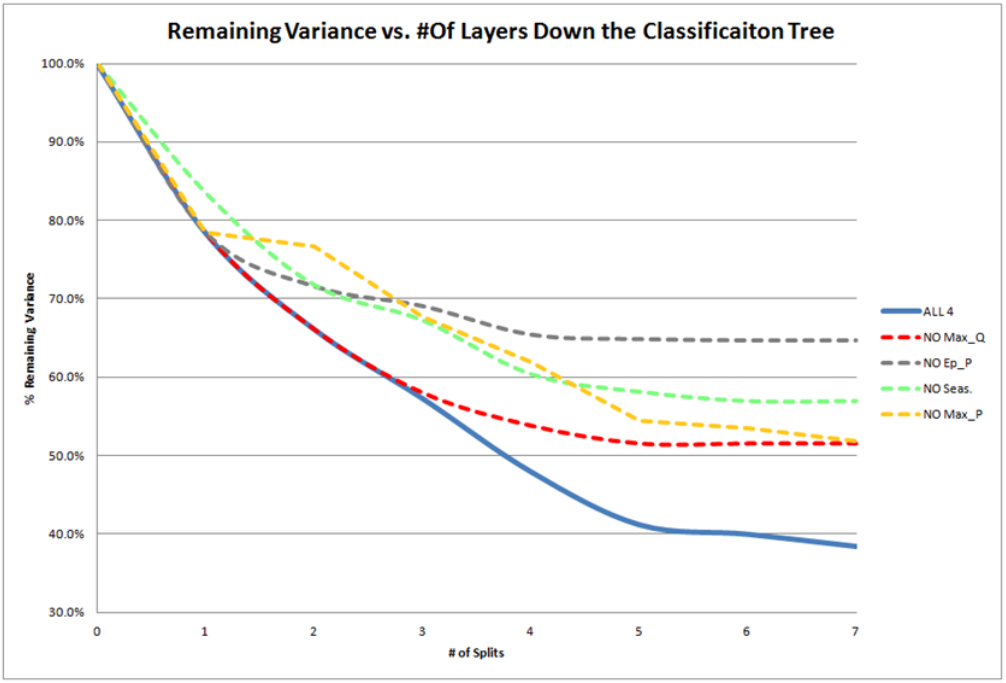
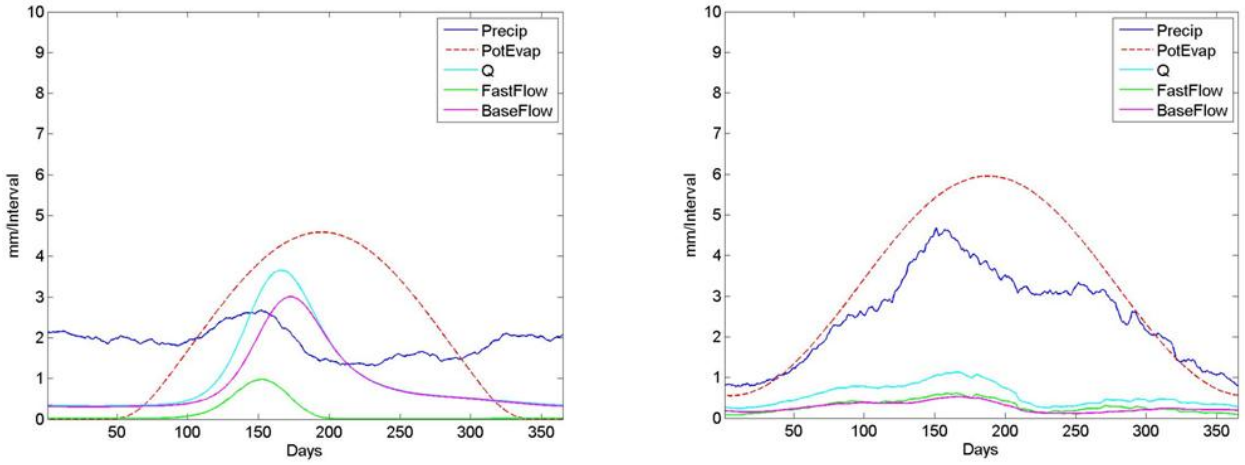


Figure A.1: Decreasing entropy down the tree –performance with each variable removed

Figures A.2 through A.5 present groups of regime curve images in which three of the four indices are very similar, but yet the holistic nature of the catchments differ dramatically, thus verifying the importance of the one index that differs.

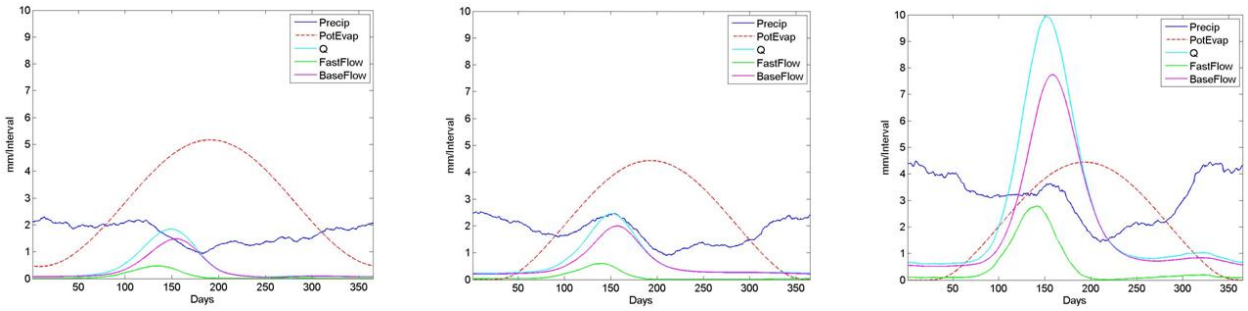
The two catchments in figures A.2a. and A.2b. are very similar in terms of aridity index, and their peak days for rainfall/runoff. However, the mountainous catchment in Montana (A.2a.) receives precipitation quite evenly throughout the year, while the Midwestern catchment in Iowa (A.2b.) receives more rainfall during summer months. The winter precipitation in Montana forms snowpacks, leading to peak runoff in the form of melt water in early June. The catchment in Iowa sees its runoff maximized during the same week as well – but the driver is rainfall, not melting. This distinction is nicely distinguished by the variable "seasonality," yet thoroughly missed by the other three variables.



Figures A.2a. and A.2b: Different seasonality measurements
(Left) #203, Montana – Ep/P ~ 1.1, MaxDayP = 152, MaxDayQ = 165… Seasonality ~ 0.15
(Right) #91, Iowa – Ep/P ~ 1.3, MaxDayP = 151, MaxDayQ = 166…Seasonality ~ 0.39

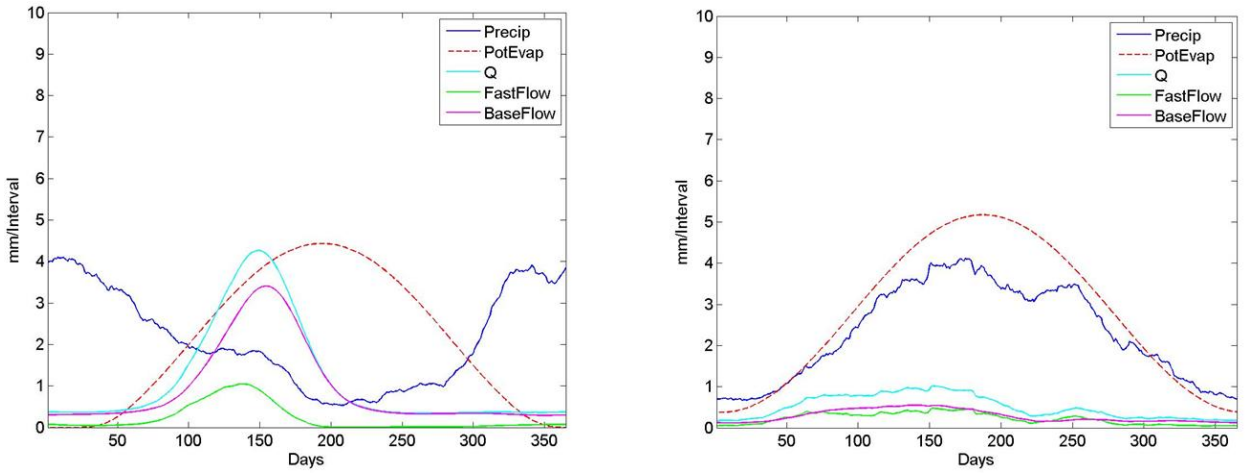
The catchments in figures A.3a through A.3c are virtually identical in terms of seasonality, as none of them display a strong seasonal signature for rainfall apart from a slightly

higher quantity of winter precipitation. All three catchments display peak rainfall within essentially one week in early January and peak runoff within the same week of late-May/early-June. Without knowledge of aridity, these catchments would almost certainly fall within the same class. However, in looking at the catchment in A.3c., the quantity of melt-driven runoff is nearly a full order of magnitude larger than the catchment in A.3a. The difference is nicely explained by the substantial differences in aridity index.



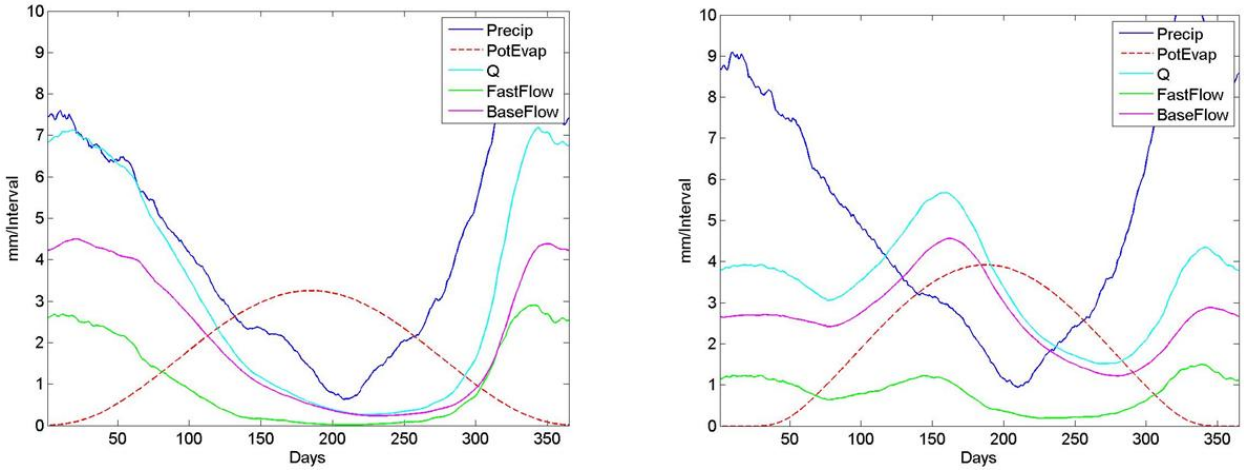
Figures A.3a. , A.3b. , and A.3c: Different aridity indices
(Left) #243, Colorado – Seasonality ~ 0.19, MaxDayP = 13, MaxDayQ = 148…Ep/P ~ 1.7
(Center) #300, Montana – Seasonality ~ 0.23, MaxDayP = 5, MaxDayQ = 151… Ep/P ~ 1.2
(Right) #342, Montana – Seasonality ~ 0.23, MaxDayP = 10, MaxDayQ = 153… Ep/P ~ 0.7

The catchments in figures A.4a and A.4b display a very similar quantity of seasonality with respect to precipitation, comparable aridity, and peak runoff during the same week. However, these catchments are distinguished by the fact that the catchment in A.4a receives its precipitation during winter, out-of-phase with respect to PE, and thus, accumulates snow which exits as melting snow months later. The catchment in A.4b, because its precipitation pattern peak is shifted almost exactly ½ year from the catchment on the left, receives precipitation in-phase with PE, and produces, despite the similar timing, dramatically less runoff.



Figures A.4a. and A.4b: Different day of peak precipitation
(Left) #162, Idaho – Seasonality ~ 0.48, Ep/P ~ 1.0, MaxDayQ = 149… MaxDayP = 8
(Right) #388, Iowa – Seasonality ~ 0.45, Ep/P ~ 1.2, MaxDayQ = 152… MaxDayP = 175

The catchments in figures A.5a and A.5b are both located in Washington. Both present significant seasonality, extremely humid climates, and seasonal precipitation that arrives out-of-phase with PE, peaking on the very same day in late November. However, these two catchments present distinctly different climates as one emits maximum runoff in December (A.5a) and the other peaks in June (A.5b). This implies a differing mechanism of runoff – A.5a receives runoff from winter rainfall that exits immediately (low residence time), shown by the Q regime curve mirroring the P regime curve, while A.5b produces runoff from winter rainfall and even more notably from spring melt, thus producing a Q regime curve that does not follow P.



Figures A.5a. and A.5b: Different day of maximum streamflow
(Left) #346, Washington – Seas. ~ 0.52, Ep/P ~ 0.38, MaxDayP = 330… MaxDayQ = 344
(Left) #392, Washington – Seas. ~ 0.49, Ep/P ~ 0.35, MaxDayP = 330… MaxDayQ = 159